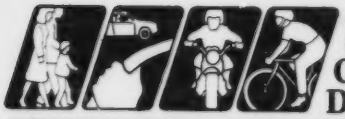
Accident Potential: An Ontario Driver Records Study Technical Report

A. Smiley
Human Factors North

E. Hauer
B. Persaud
Department of Civil Engineering
University of Toronto

L. Clifford
D. Duncan
Ministry of Transportation

April, 1991
ISBN 0-7729-5848-3
SCDO 91-116



Safety Co-ordination & Development Office
Research & Evaluation Section

Ministry of Transportation
Ontario

ABSTRACT

The current demerit point system in use in Ontario allocates points to offences on the basis of the perceived seriousness of the offence. The goal of the work reported here was to allocate points to offences with a different purpose in mind. The purpose here was to use a driver's record of convictions and accidents in order to predict, as well as possible, which drivers, based on their past record, are most likely to have an accident in the near future.

A sample of 827,955 records of Ontario drivers containing information about age, gender, convictions, accidents, demerit points and suspensions, for 1981-1984 was examined. On this basis 16 new models to estimate a driver's accident potential were formulated. Some models used age and gender information, others did not. Some models assigned different weights to each of the conviction categories; in other models all convictions were weighted equally (equivalent to 1 point per conviction). In some models accident data were used; in other models, not. In contrast to the current system, information on all types of convictions, not just moving convictions, was used. It turns out that these new models are better at predicting a driver's accident potential than the current demerit point system.

The performance of the new models is described by means of two measures. The first measure is the number of accidents in period 2 (1983-1984) based on the driver's record of convictions and accidents in period 1 (1981-1982). For example, if the 10,000 drivers with the **most demerit points** in period 1 are selected, 1452 have accidents in period 2; if the 10,000 drivers with the **most accidents** in period 1 are selected, 1828 have accidents in period 2. Alternatively, if the 10,000 drivers are selected on the basis of the **best new model**, which uses age, gender, accidents and separate weights for 14 offence categories, 2116 have accidents in period 2.

The second measure of performance is in terms of "hits" and "false alarms". "Hits" are drivers who have an accident potential in excess of 4 times the population average. For example, if the 10,000 drivers with the highest estimated accident potential are selected by the best model, 3757 of these are are expected to be "hits". "False alarms" are drivers who have an accident potential which is below the population average. Of the same 10,000, one should expect 792 to be "false alarms". The best model identifies approximately twice as many high accident potential drivers as the current demerit point system. Even the simplest model, which uses total convictions as the only variable, predicts 50% more high accident potential drivers than the current system.

ABRÉGÉ

L'actuel système de points d'inaptitude en usage en Ontario donne des points aux infractions en fonction de la gravité qu'on leur attribue. Dans le travail qui nous occupe ici, il s'agissait de donner des points aux infractions dans un but différent, plus précisément, de se baser sur le dossier des condamnations et accidents des conducteurs pour prédire le mieux possible lesquels ont le plus de chances d'avoir un accident dans un avenir proche.

On a donc examiné un échantillon de 827 955 dossiers de conducteurs ontariens, contenant des renseignements sur l'âge, le sexe, les condamnations, les accidents, les points d'inaptitude et les suspensions de permis de 1981 à 1984. À partir de ces données, on a établi 16 nouveaux modèles censés évaluer les chances d'accident d'un conducteur. Certains modèles intégraient les données sur l'âge et le sexe, d'autres pas. Certains modèles pondéraient les diverses catégories de condamnations et d'autres les traitaient toutes sur le même pied (un point par condamnation, quelle que soit la nature du délit). Certains modèles intégraient les données sur les accidents et d'autres pas. Contrairement au système actuel, ces modèles tenaient compte de tous les types de condamnations, et non uniquement de celles portant sur des infractions commises par les conducteurs de véhicules en marche. Il s'est avéré que ces nouveaux modèles arrivent à mieux prédire les chances d'accident d'un conducteur que l'actuel système de points d'inaptitude.

Les résultats des nouveaux modèles apparaissent dans deux mesures. La première est le nombre d'accidents pendant la période 2 (1983-1984) en fonction du dossier des condamnations et accidents du conducteur pendant la période 1 (1981-1982). Par exemple, si l'on retient les 10 000 conducteurs qui ont eu le plus de points d'inaptitude pendant la période 1, on trouve que 1 452 d'entre eux ont des accidents pendant la période 2; si l'on retient les 10 000 conducteurs qui ont eu le plus d'accidents pendant la période 1, on trouve que 1 828 d'entre eux ont des accidents pendant la période 2. Selon l'autre méthode, si la sélection des 10 000 conducteurs s'effectue selon le meilleur nouveau modèle, basé sur l'âge, le sexe, les accidents et 14 catégories d'infractions de poids différents, on trouve que 2 116 d'entre eux ont des accidents pendant la période 2.

La deuxième mesure classe les conducteurs en «récidivistes» et «fausses alertes». Les récidivistes sont les conducteurs qui ont des chances d'accident supérieures à quatre fois celles de la population moyenne. Par exemple, si les 10 000 conducteurs qui ont le plus de chances d'accident sont sélectionnés par le meilleur modèle, 3 757 d'entre eux sont classés comme «récidivistes». Les «fausses alertes» sont les conducteurs qui ont des chances d'accident inférieures à celles de la population moyenne. Sur les 10 000 conducteurs cités ci-dessus, on devrait en trouver 792. Le meilleur modèle permet d'identifier environ deux fois plus de conducteurs ayant des chances élevées d'accident que l'actuel système de points d'inaptitude. Même le modèle le plus simple, dont la seule variable est le nombre total de condamnations, identifie 50 % de plus de conducteurs ayant des chances élevées d'accident que l'actuel système.

TABLE OF CONTENTS

1.0	INTRODUCTION	Page 1
2.0	DRIVER RECORD SAMPLE	1
3.0	PREPARATION FOR ANALYSIS	4
3.1	Checks for Representativeness of the Sample	4
3.2	Consistency Checks	4
3.3	Accident Related Convictions, Multiple Convictions, and At-Fault Accidents	7
3.4	Exploratory Analysis	7
3.5	Selection of Conviction Categories	10
3.6	Discussion of Conviction Categories	17
3.7	Age and Gender Categories	20
3.8	Exclusion of Drivers With Suspended Licenses	20
4.0	ANALYSIS AND RESULTS	22
4.1	Schemes and Variants Examined	22
4.2	What is a "Model" and how does it "Work"	23
4.3	On the Estimation of "Weights" and "Diversity"	25
4.4	The Accident Potential of Ontario Drivers	28
4.5	The Performance of the Current Demerit Point System and of the New Models	30
5.0	SUMMARY AND CONCLUSIONS	44
APPENDIX 1	Recalculation of the Estimated m Associated With Each Offence Type According to a Standardized Age-Gender Distribution	A1.1
APPENDIX 2	Number of Convictions in Each of the Final 14 Conviction Categories	A2.1
APPENDIX 3	Model Weights for the Alternate Demerit Point Schemes	A3.1
APPENDIX 4	Theoretical Framework	A4.1

Figures

Figure 1	4 Year Convictions vs. 4 Year Accidents	Page 9
Figure 2	2 Year Convictions vs. 2 Year Accidents	11
Figure 3	Speed Limit Violations, 60 km/h	14
Figure 4	Speed Limit Violations, 100 km/h	15
Figure 5	3 Yr m's For Drivers With 1 Conviction	19
Figure 6	Period 2 Accidents By Age and Sex	21
Figure 7	Estimated Coefficients, Model A2	27
Figure 8	Distribution of Accident Potential in the Ontario Driving Population	29
Figure 9	Distribution of Drivers with $m > 0.10$	31
Figure 10	Distribution of Drivers with $m > 0.20$	32
Figure 11	Distribution of Drivers with $m > 0.30$	33
Figure 12	Hypothetical Curve of Distribution of Expected Numbers of Accidents	37
Figure 13	Current DP vs. Schemes A4, B4, C4, and D4	42
Figure 14	Runs A4, B4, C4, D4	43

Tables

Table 1	Summary of Data Extracted from MTO driver records	2-3
Table 2	Description of Data in On-Line Master File	5
Table 3	Comparison of Sample with Population	6
Table 4	Multiple Convictions (Same Day) in Sample of 7811 Drivers with Ranking According to Reason for Apprehension	8
Table 5	Accidents for Drivers with 1 Conviction in 1 Year	12
Table 6	Accidents for Drivers with 1 Speeding Conviction in 1 Year	16
Table 7	3-Year Mean Accidents for Drivers with 1 Conviction in 1 Year (Based on Final Conviction Categories)	18
Table 8	Variables Used for Regression Runs	24
Table 9	Coefficients for Model A2	26
Table 10	Accidents Per Year Recorded by Drivers Selected by Various Models	35
Table 11	Figures of Merit for Models A4, B4	39
Table 12	Figures of Merit for 10,000 Drivers With Highest m's (for each model)	41

1.0 INTRODUCTION

The current demerit point system in use in Ontario allocates points to offenses on the basis of the perceived seriousness of the offense. An offense is considered serious if it is associated with a relatively large chance of precipitating an accident. This is why a non-moving violation, such as not having a trailer permit, receives no points, while running a red light receives a lot of points. The goal of the work described in this report was to allocate points to offenses with a different purpose in mind. The purpose here was to use a driver's record of convictions and accidents in order to estimate as well as possible which drivers, based on their past record, are most likely to have an accident in the near future. This is done by calculating for each driver how many accidents per year he or she is likely to have, on the average.

Because a person's "accident potential" can only be indirectly estimated (not directly measured) and because, mercifully, for any one driver it is rare to be involved in an accident, the accuracy with which we can estimate the accident potential of a driver is bound to be severely circumscribed. Thus, our aim is not only to produce for each driver an estimate of his or her accident potential but also to say how accurate our estimate is.

This is the kind of knowledge which might then be used in the determination of post-licensing control action. Thus, while a non-moving violation may not be a threat to traffic safety, such a conviction on a driver's record may be an important clue about that person's likelihood of future accidents.

2.0 DRIVER RECORD SAMPLE

The analysis examined driver records over a recent four-year period, January 1981 to December, 1984. Convictions and accidents during the first two years of the record were used as the basis for estimating a person's accident potential in the second two-year period.

To be included in the sample, drivers had to have been licensed to drive in Ontario throughout the four-year period. To this end, a number of criteria were used in the selection of driver records. These criteria were as follows. First, drivers not licensed at the beginning of the study period were excluded. Second, drivers who might have moved out of Ontario during the four year period were excluded. These exclusions were achieved by only selecting drivers who renewed their drivers' license during 1985, with the exception of those who were 16 or 17 in 1981. This exception was made because such a large percentage of drivers become licensed at 16 and 17, that the majority of drivers of this age would be renewing their licenses in 1984 (3 years after the original licensing) and would have been excluded from the sample. Because drivers had to have been licensed in 1987 to be included in this study, no drivers on the new probationary license system were included.

Of the 5.5 million Ontario drivers, 827,995 met these criteria and were included in the sample. Data extracted from the MTO driver records included the following information: age, gender, and for each conviction, type, date, and demerit points assigned; for each accident, degree of severity, date, and for each suspension, type and time period. Details are shown in Table 1.

TABLE 1 Summary of data extracted from MTO driver records

1. Root Segment (33 characters)

Field #	Width	Description
1.1	15	License Number
1.2	3	License Class
1.3	3	Total Demerit Points at Time of "Dump"
1.4	6	Original Date of License (YY-MM-DD)
1.5	6	Date of Last Renewal (YY-MM-DD)

2. Conviction Segment (Space for 75 convictions at 22 characters per conviction

Field #	Width	Description
2.1	6	Conviction Offence Date (YY-MM-DD)
2.2	5	" " Code
2.3	3	Recorded Speed for Speeding Convictions
2.4	3	Speed Limit " " "
2.5	3	Conviction Demerit Points
2.6	1	Code = 1 for Personal Injury, = 2 for Property Damage, 3 for Fatality, associated with conviction.
2.7	1	Code = 1 if partial points applied, Ø otherwise.

3. Suspension Segment (Space for 2Ø Suspensions at 15 characters per suspension

Field #	Width	Description
3.1	6	Suspension Effective Date (YY-MM-DD)
3.2	2	Suspension Type
3.3	6	Suspension Expiration Date (YY-MM-DD)
3.4	1	Code =1 if suspension cancelled on appeal

4. Collision Segment (Space for 25 collisions at 1Ø characters per collision

Field #	Width	Description
4.1	6	Collision Effective Date (YY-MM-DD)
4.2	1	Collision Nature (Severity)
4.3	2	Apparent Driver Action
4.4	1	Driver Condition

Table 1: (ctd.)

5. Cancellation Segment (Space for 1Ø cancellations at 6 characters per cancellation

Field #	Width	Description
5.1	6	Cancellation Effective Date (YY-MM-DD)

6. Medical Condition (Space for 1Ø at 1 character each)

Field #	Width	Description
6.1	1	Code 1 if code 43 (drug abuse) is flagged, 2 if code 6Ø (Alcoholism - no seizure) is flagged 3 if code 61 (Alcoholism - seizure) is flagged

7. Demerit Point History (Space for Field 7.1 plus 15Ø adjustments -- 7.2 to 7.4 -- at 9 characters each)

Field #	Width	Description
7.1	4	Total Demerits points at January 1, 1981.
7.2	4	Month and Year of Adjustment (YY-MM)
7.3	4	Change to Demerit Point Total
7.4	1	Code 1 for Reinstatement

Of the 827,995 drivers, 311,691 had no convictions or accidents during 1981-1984. To save on storage space, these records were kept in a separate file. Of the 516,304 drivers who had at least one conviction and/or accident during 1981-1984, records for approximately 400,000 were stored in an on-line file. The layout of this file is shown in Table 2.

The processing of this amount of data can be costly. For some purposes, smaller data sets are sufficient. With this in mind, three sets of driver records were assembled: a full set of records, and smaller samples of every 40th record and every 400th record.

3.0 PREPARATION FOR ANALYSIS

Making sense of large data sets requires careful preparation. The several preparatory activities are described below.

3.1 Checks for Representativeness of the Sample

Several tests were conducted to check whether the characteristics of a smaller sample (1 in 40 records) were similar to known (published) characteristics of the Ontario driving population. In terms of percent of male and female drivers, percent of accidents involving males, and accident involvement per driver/year the sample data were found to be reasonably representative of the Ontario driving population (See Table 3).

3.2 Consistency Checks

Driver record data were then checked for consistency. A number of inconsistencies were noted. The first inconsistency apparent in the data was that while some offenses are logically associated with the occurrence of an accident, the count of such offenses exceeded the count of related accidents. Thus, for example, one would expect that if a driver were convicted of failing to remain at the scene of an accident, he or she would also show an accident associated with the offense on that same day. However, this is clearly not so; in the small sample there were 45 such convictions, but only 28 accidents for drivers convicted of this offense. There is a similar discrepancy for convictions for "following too close" which apparently arise mainly from accidents. These discrepancies suggest that many accidents related to convictions have not been entered in the driver record file. It is possible that this omission might apply to accidents in general. One reason for this may be due to the fact that records contain only reports of "reportable" accidents whereas charges apply to any.

A second inconsistency was the considerable year-to-year fluctuation in the number of convictions for certain offenses. For example, the number of convictions for driving without a permit increased from 0 in 1981 to 68 in 1984. There were seven other conviction types with similar inconsistencies from year-to-year. The reason for these inconsistencies is not clear, but they could be related to changes in enforcement practices or to modifications related to regulations.

TABLE 2: Description of data in on-line master file

Field #	Width	Description
Record 1: Basic record (Fixed record length = 42)		
1A	2	Birth Year
1B	1	Sex: Code 1 for male, 2 for female
1C	3	License class(es)
1D	3	Demerit Point Total at time of "dump"
1E	2	Number of convictions, 1981-1984, n1
1F	2	Number of Accidents, 1981-1984, n2
1G	2	Number of demerit point adjustments, 1981-84, n3
1H	8x3	Suspension days unpaid fine, 1981 " " other , 1981 " " unpaid fine, 1982 . . . . Suspension days other , 1984
1I	1	Code 1 if any flags for drugs or alcohol
1G	2	Demerit Point Total at January 1, 1981
Record 2: Conviction Record (Variable - space for n1 convictions)		
2A	1	Year (1981=1, and so on)
2B	5	Conviction Code
2C	3	Speed Limit (for speeding convictions)
2D	3	Recorded Speed for Speeding Conviction
2E	2	Demerit Points assigned
2F	1	Code 1 for non-accident related, single conviction 2 for non-acc. related, primary of multiple conv. 3 for non-acc. related, secondary of multiple " . 4 for accident related, single conviction 5 for accident related, primary of multiple conv. 6 for " " , secondary of " " .
Record 3: Accident Record (Variable -- space for n2 accidents)		
3A	1	Year (1981=1, and so on)
3B	2	Severity code
3C	3	Apparent Driver Action Code
3D	1	Driver Condition Code
Record 4: Record of demerit point adjustments (Variable record length -- space for n3 adjustments)		
4A	2	Month # from Jan. 1981, of adjustment
4B	1	Sign of adjustment: + = points added
4C	2	Demerit Point change
4D	1	Code 1 if reinstatement after suspension

TABLE 3: Comparison of sample with population*

Characteristic	Sample Value	Population value
% male drivers	53.4%	56%
% female drivers	46.6%	44%
% of accidents involving males	72.2%	72.9%
% of accidents involving females	27.8%	27.1%
Involvements per driver/year	Ø.Ø64	Ø.Ø57
Proportion fatal & injury involvements (all accidents)	Ø.35	Ø.35

* - Population values are estimated from the Ontario MTO 1985 Annual Accident report and are for a comparable age distribution.

The third inconsistency involved the original licensing date. This variable was found to be unreliable. Some of the dates were missing and none of the the original dates were earlier than 1966 although a substantial number of drivers were obviously licensed before that year. This meant that the effect of experience on a driver's accident record could not be examined. This is a result of the development of the computer system in 1966 and of changes of residence and lapses of renewal that affect the recording of the original date.

3.3 Accident Related Convictions, Multiple Convictions and At-Fault Accidents

We anticipated that some models for the estimation of accident potential would make use of the driver's accident record. To avoid double-counting, convictions which were the consequence of accident involvement had to be identified. For the purpose of later analysis, we identified a conviction as being accident-related if it occurred on the same date as an accident. In a sample of 16,000 drivers we found 8940 convictions in two years. Of these, 612 (7%) were for violations which the driver incurred on the same day as an accident.

If several convictions were recorded on the same date, they were assumed to be related to a single apprehension. We think that only the "primary" conviction should feature in the statistical model; inclusion of secondary convictions could obscure the relationship between cause and effect.

Each offense type was assigned a rank, as shown in Table 4 according to how likely it was to be the cause of an apprehension. Thus, for example, speeding was given the highest rank. For multiple convictions, the one for the offense with the highest rank was called the primary conviction, while the others were designated as secondary convictions. In case of a tie, the first conviction recorded was designated as the primary one. This logic is somewhat arbitrary, but analysis of a data sample showed that only about 4% of all apprehensions result in multiple convictions. In addition, some 55% of the multiple convictions involved speeding - an obvious primary offense.

The information in Field 3C (See Table 2) - Apparent Driver Action - was used subsequently in the determination of whether or not an accident was "at-fault". All accidents with apparent driver action codes other than "driving properly", were considered at-fault accidents. These accidents included some 12% of involvements for which apparent driver action was either "not known" or "other". The decision to place these involvements in the "at-fault" category was made after discussions with police. Also, since "driving properly" already accounted for about 50% of involvements, it was logical to assume that the rest of the involvements were "at-fault".

3.4 Exploratory Analysis

The sample of every 40th record was used to perform some exploratory analyses. Figure 1 depicts the relationship between accidents and convictions in four years in a sample of nearly 16,000 drivers. It shows that drivers with more convictions tend to have more accidents. Thus, for example, the 189 drivers with 6 convictions in four years had, on the average, eight times as many accidents (line 1) as drivers with no convictions in these four years.

TABLE 4:

Multiple convictions (same day) in sample of 7811 drivers with ranking according to reason for apprehension:

1 = most likely; 4 = unlikely

Offence	Code	No. in Sample
Rank 1		
Dr. m/v, no valid plates	10034	1
No plate, operate m/v	10063	3
Imp./insuff./ drvg. lamps	10420	9
Imp. no. plate lamp	10530	6
Excess fumes/smoke	10870	2
Unnecess. noise	10880	7
Horn/bell violation	10890	1
Speeding -- km in --km zn.	11355	211
Careless driving	11490	22
Fail to stop, intersec.	11580	25
Fail to yield ROW at stop	11590	2
Fail yield ROW at yield	11610	1
Fail to yield ROW, Pr. rd.	11620	2
PXO viol. oth. side of rd.	11650	1
Pass at Ped. crossing	11665	2
Imp. RT at intersection	11693	1
Improper left turn	11713	2
Imp. LT at intersection	11723	3
Improper left turn	11733	1
" " from parked pos.	11760	3
Disobey red light	11860	26
" amber "	11870	2
Disob. green arrow	11900	1
Prohibited turn	11920	10
Disobey red light	11925	2
Fail to share road	11940	3
Imp. pass.- appr. traffic	12010	2
Driving left of center	12040	2
Pass on right/off r'way	12100	1
Wrong-way,one-way street	12110	2
Imp. driving, div. h'way	12120	4
Disobey off.signs,div.h'way	12140	1
Fail to stop for emer/veh.	12200	1
Imp. park., Interf. w.traf	12430	1
Prohibited turn	80020	1
By-law, proh. Turn	80110	1
Rank 2		
Imp. rear lamp on trl.	10570	1
View obstr., sign/poster	10800	1
Rear window obstructed	10850	1
Def/imp/no muffler	10860	4
F.T.C., motor vehicle	12180	4
Fail to rep. accident	12710	16
Fail remain at scene accid.	12730	10
Fail rep. dam. to hy. prop.	12760	2
Dang. driving CCC	70060	1
Fail rem at accid. CCC	70090	6
Danger. driving CCC	70120	2
Impaired driving C.C.C.	70174	23
Fail/ref. breath test CCC	70185	1
Fail or ref. breath sam.CCC	70212	19
Driving with >80 mgs.alcoh.	70214	15
Rank 3		
Imp. signal devices	10610	1
No safe helmet, motor cy.	11090	3
Fail/imp. use of Seat Belt	11097	85
Insecure load	11260	2
Excess veh. width	11310	1
Excess len. veh-comb.	11320	1
Excess veh. height	11350	1
Crowding driver seat	12240	1
Rank 4		
Dr. m/v,no valid permit	10033	19
No valid. tag on plates	10035	7
Draw trailer, no permit	10037	3
fail surr. m/v permit	10038	7
" " trl. "	10039	1
fail not. ch. address	10050	3
Fail apply for permit	10071	2
Imp. plates on veh.	10130	3
No driver's license	10190	39
Driv., cond. proh.	10193	2
Fail to prod. dr. license	10210	45
No lic., fail to ID self	10214	1
Pos. canc/susp. license	10222	2
Unlaw. pos. license	10320	1
Dr. lic. susp. H.T.A.	10332	31
Def. brakes, contr. to reg.	10680	1
Def./imp tires	10750	2
Refuse exam. unsafe veh.	10980	1
Fail to carry/prod.permit	13210	3
More than 1 license	13270	1
No insurance	20010	2
Op. m/v, no insurance	20011	5
False statem. re. insurance	20030	1
Fail to have insurance card	20050	28
" " carry " "	20060	2
" " " evid. of ins.	20070	1
Dealer plate infraction	90250	1
Learner unaccompanied	90700	1
" " fail not. addr. ch.	90810	6

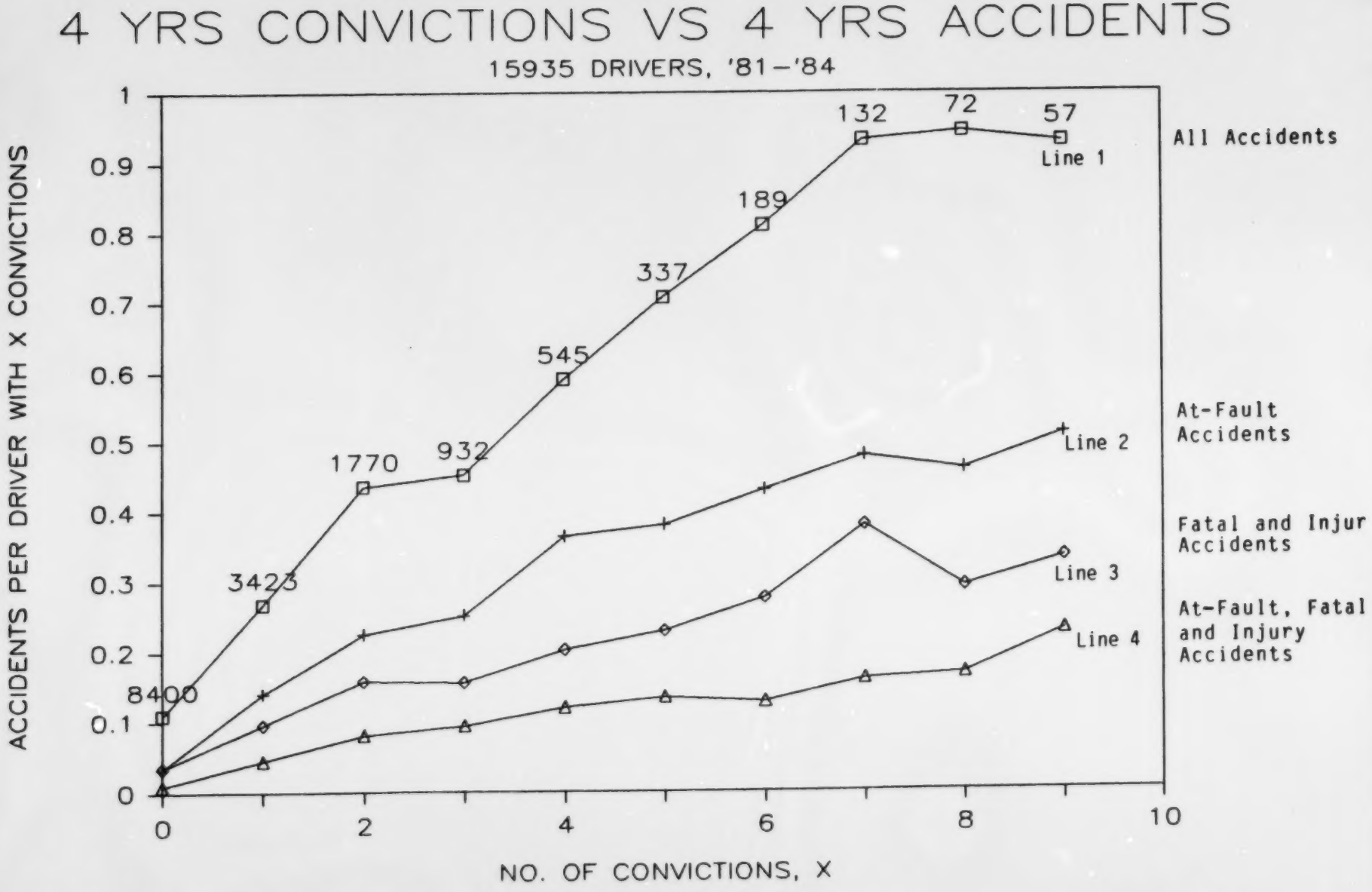

4 YRS CONVICTIONS VS 4 YRS ACCIDENTS
15935 DRIVERS, '81–'84
ACCIDENTS PER DRIVER WITH X CONVICTIONS
NO. OF CONVICTIONS, X
All Accidents
At-Fault Accidents
Fatal and Injur Accidents
At-Fault, Fatal and Injury Accidents
Line 1
Line 2
Line 3
Line 4
8400
3423
1770
932
545
337
189
132
72
57
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
2
4
6
8
10

Figure 1

(Convictions which were incurred on the date of an accident were eliminated from consideration). The same pattern prevails for at-fault accidents (line 2), fatal and injury accidents (line 3) and at-fault, fatal and injury accidents (line 4). Figure 2 shows similar results. However, this time the relationship between convictions in the first two year period and the average number of accidents in the subsequent two-year period is plotted.

These two figures leave no doubt that convictions and accidents are correlated. However, correlation should not be taken to indicate causality. It is, in fact, very likely that the number of accidents in which a driver is involved and the number of convictions which a driver incurs are both related to the distance driven. Thus, the observed statistical correlation reflects a hidden common link, exposure, about which we have no information.

It deserves noting that about 2/3 of all drivers have no convictions in a two-year period. We are interested mainly in the accident potential of drivers who have convictions. In addition, the difficulties of computing are greatly reduced if only that 1/3 who had one or more convictions are used for data. It is for these reasons that the core of the analysis described in Section 4 used the records of drivers who in the first two years had one or more convictions.

3.5 Selection of Conviction Categories

A preliminary analysis of a sample of about 8000 drivers indicated that, over the four-year period 1981-1984, these drivers were convicted for approximately 200 different traffic offenses. Speeding accounted for some 60% of non-accident related convictions; seat-belt offenses accounted for about 10%, and failure to stop at an intersection for about 5%. Most of the other offense types had very few convictions. It was obviously impractical to estimate how much these "leaner" convictions added to a driver's accident potential. It was therefore necessary to aggregate offenses with few convictions into larger groups. (It should be noted that the same approach is already present in the current demerit point-system. There are essentially 8 categories of offenses, that is, those assigned 0,1,2,3,4,5,6 and 7 demerit points. For example, all the non-moving violations as well as some of the moving ones fall into the 0 point category.)

Determining Conviction Categories: Step 1

The clustering of all possible offense types into a manageable number of groups was accomplished in several steps. The first step was to combine the offense types which are quite similar in nature into a smaller number of categories. A total of 45 such categories were established in our discussions with MTO in April 1987. These consisted of 38 categories of moving violations, 3 for vehicle-related offenses, and 4 for non-moving, administrative violations. (See Table 5.)

This first round of clustering was based on judgment of similarity. In the second round, groups of offenses were identified which were similar in their contribution to the average number of accidents of a driver. To obtain results which are unambiguous and free of confounding, the only records used

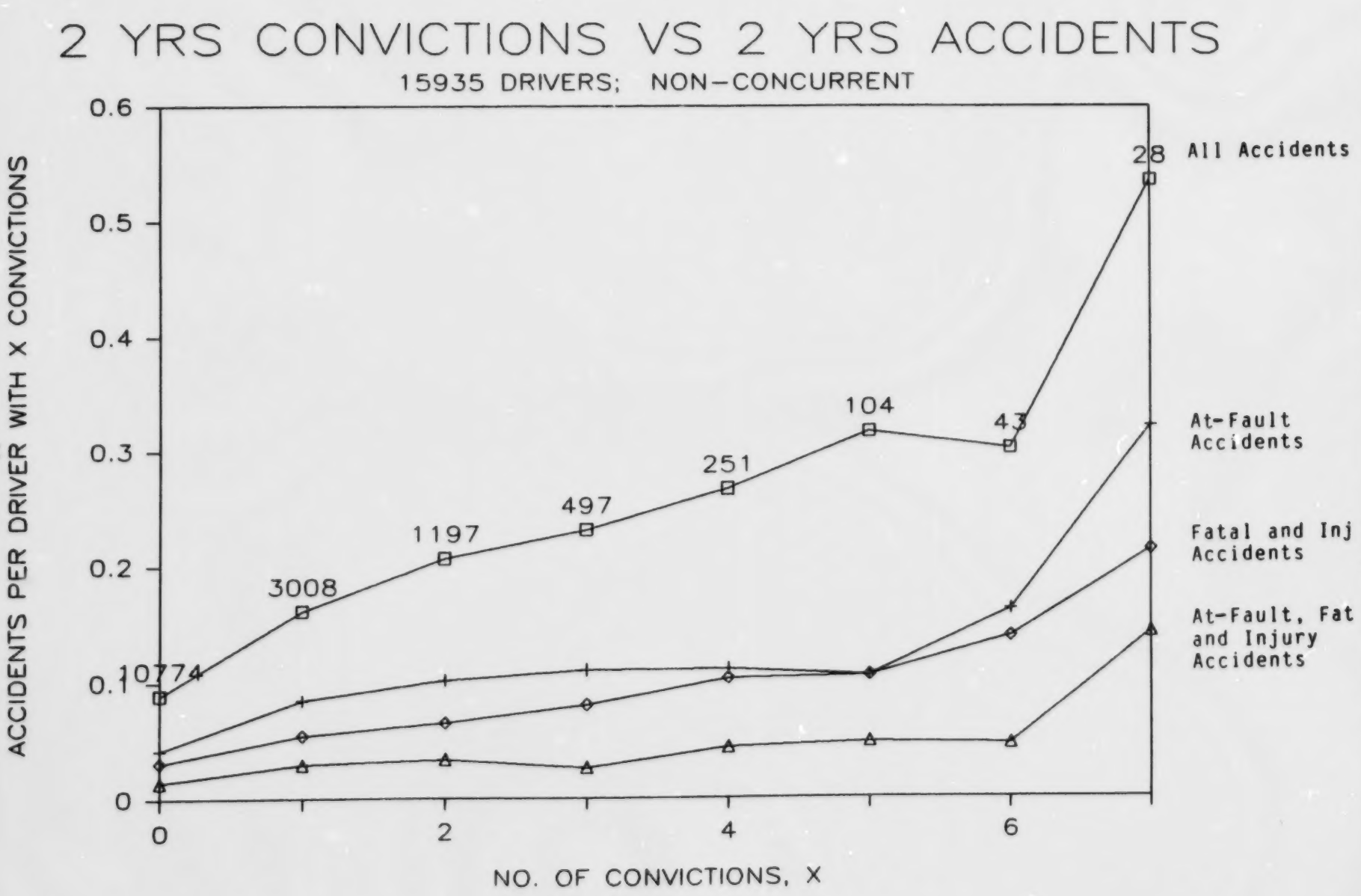

2 YRS CONVICTIONS VS 2 YRS ACCIDENTS
15935 DRIVERS; NON-CONCURRENT
ACCIDENTS PER DRIVER WITH X CONVICTIONS
0.6
0.5
0.4
0.3
0.2
0.1
0
10774
3008
1197
497
251
104
43
28
All Accidents
At-Fault Accidents
Fatal and Inj Accidents
At-Fault, Fat and Injury Accidents
0
2
4
6
NO. OF CONVICTIONS, X

Figure 2

TABLE 5: Accidents for Drivers with 1 Conviction in 1 Year

Category	Brief Description	No. of Drivers	3 yr. Accs.	Weighted Mean	95% Limits Upper	Lower
n1	Minor neglect, licenses, permits	6495	2918	0.434	0.445	0.422
n2	Neglect, insurance, permits, etc.	1589	719	0.414	0.438	0.392
n3	License suspended, HTA	874	454	0.424	0.456	0.394
n4	Learners	34	18	0.343	0.502	0.212
v1	Minor veh.; lamps, noise	2954	1498	0.468	0.485	0.451
v2	Brakes, tires, unsafe vehicle	946	451	0.400	0.430	0.371
v3	Comm. veh; size & weights	503	369	0.542	0.583	0.500
m1	Seat belt	12337	4858	0.376	0.384	0.368
m2	Speeding	173592	55211	0.319	0.321	0.317
m3	Careless driving	902	342	0.327	0.357	0.299
m4	Slow driving	45	11	0.119	0.237	0.055
m8	STOP sign, ROW violations	14024	3935	0.288	0.295	0.281
m9	PXO violations	1237	355	0.296	0.320	0.272
m10	Turns violations; right, left, U	18231	4942	0.283	0.289	0.277
m11	Unsafe move; open door	1649	542	0.334	0.355	0.312
m13	Disobey red light	13731	4270	0.313	0.321	0.306
m14	Amber light	3453	982	0.285	0.299	0.271
m15	Advance green	274	73	0.265	0.317	0.218
m16	Fail to share road	170	62	0.303	0.372	0.242
m17	Passing violations	1305	459	0.327	0.351	0.303
m18	Wrong-way one way street	1582	458	0.284	0.306	0.264
m19	Improper driving divided h'way	2599	900	0.361	0.379	0.344
m20	F.T.C.	934	337	0.344	0.374	0.316
m21	Emerg. veh., school x'ing	48	15	0.159	0.280	0.084
m22	R/R crossing violations	95	35	0.314	0.408	0.233
m24	Headlight beam not lowered	225	71	0.260	0.318	0.209
m25	Improper parking	145	77	0.407	0.484	0.334
m26	Fail stop for school bus	604	133	0.281	0.317	0.249
m28	Disobey traffic signs	1650	529	0.322	0.344	0.301
m29	Fail report accident	224	73	0.266	0.324	0.215
m30	Fail remain at scene	236	75	0.315	0.373	0.261
m32	Dangerous driving C.C.C.	5	2	0.010	0.421	0.000
m33	Fail remain at accident C.C.C.	66	39	0.377	0.491	0.274
m34	Dangerous driving C.C.C.	89	46	0.281	0.376	0.202
m35	Impaired driving C.C.C.	2381	1040	0.443	0.462	0.424
m36	Fail/refuse breath test C.C.C.	94	37	0.218	0.306	0.149
m37	Fail or ref. breath test C.C.C.	120	50	0.314	0.397	0.241
m38	Driving with >80 mgs. alcohol	3676	1502	0.386	0.401	0.371
m41	Crowding driver seat	120	36	0.280	0.362	0.211
m44	Radar device in vehicle	68	27	0.251	0.359	0.166
m45	No safe helmet, motorcycle	96	47	0.266	0.357	0.191
m46	Fail to signal to stop	18	6	0.080	0.277	0.019
m47	FTC, commercial vehicle	67	35	0.362	0.475	0.262
m48	Fail to stop for police officer	12	6	0.086	0.340	0.016

were those for drivers who in any one of the four years had a single conviction. The results given in Table 5 are based on the entire sample. Thus, for example, there were 12337 drivers who in one year were convicted for not wearing a seat belt (offense m1) and had no other convictions in that year. During the remaining three years these drivers recorded 4858 accidents for an average of 0.394 accidents per driver. To eliminate any bias due to differences in the age-gender distribution which might be associated with specific offenses, all averages were recalculated for a "standard population" (for more detail, see Appendix 1). The "standard population" used had an age-gender composition which was characteristic of those drivers who had exactly one conviction of any kind in one of the four years. This is why in the "Weighted Mean" column the average number of accidents in the remaining three years associated with this seat belt offense is listed as 0.376. Similarly, the 173,592 drivers who had only a single speeding conviction (m2) in some year have an adjusted average of 0.319 accidents in three years. The last two columns give 95% confidence limit for the weighted mean.

Categorization of Speeding Offenses: The current point allocation system establishes four levels of severity according to how much the speed limit is exceeded. However, exceeding the speed limit by 20 km/h, whether in a 40 km/h or a 100 km/h speed zone, carries the same number of demerit points. Since there were many speeding convictions in the data base, the question of how the average number of accidents depends on the speed limit and on the extent to which the speed limit is exceeded was examined.

Figures 3 and 4 show the frequency of violations for various speeds over the limit. It seems clear that there is a tendency for "speeds over limit" to be rounded to multiples of 5 km/h. However, the largest numbers of violations are for 14 km/h over the speed limit. Since demerit points are given for 15 km/h or more over the speed limit, it appears that either drivers consciously drive just below this number or, more likely, that there is some mitigation by police officers for speeds just beyond 14 km/h over the speed limit (perhaps to account for inaccuracies in radar gun and speedometer readings). A similar situation exists, though less pronounced, for 29 km/h over the limit, just below the 30 km/h at which demerit points are increased from 3 to 6.

As a result of the anomaly noted above, dividing speeding convictions for the purpose of determining the relative seriousness of various "speeds over limit" can produce misleading results. In a division based, say, on the current system, most convictions in the 0-14 km/h over the limit category should really belong to the seemingly more serious 15-20 km/h over limit category. Similarly, the 15-29 km/h category may contain many convictions that belong in the 30 km/h or more over the limit category. Thus regression weights for each category might wrongly indicate that the 3 categories are similar in terms of seriousness.

We have examined in detail whether drivers who have in some year exactly one speeding conviction of given gravity differ in their average number of accidents during the remaining three years. The results are given in Table 6. It turns out that there is no systematic or practically significant difference between the average number of accidents either by speed limit zone or by the amount by which the speed limit is exceeded, or by the combination of both criteria.

FIGURE 3

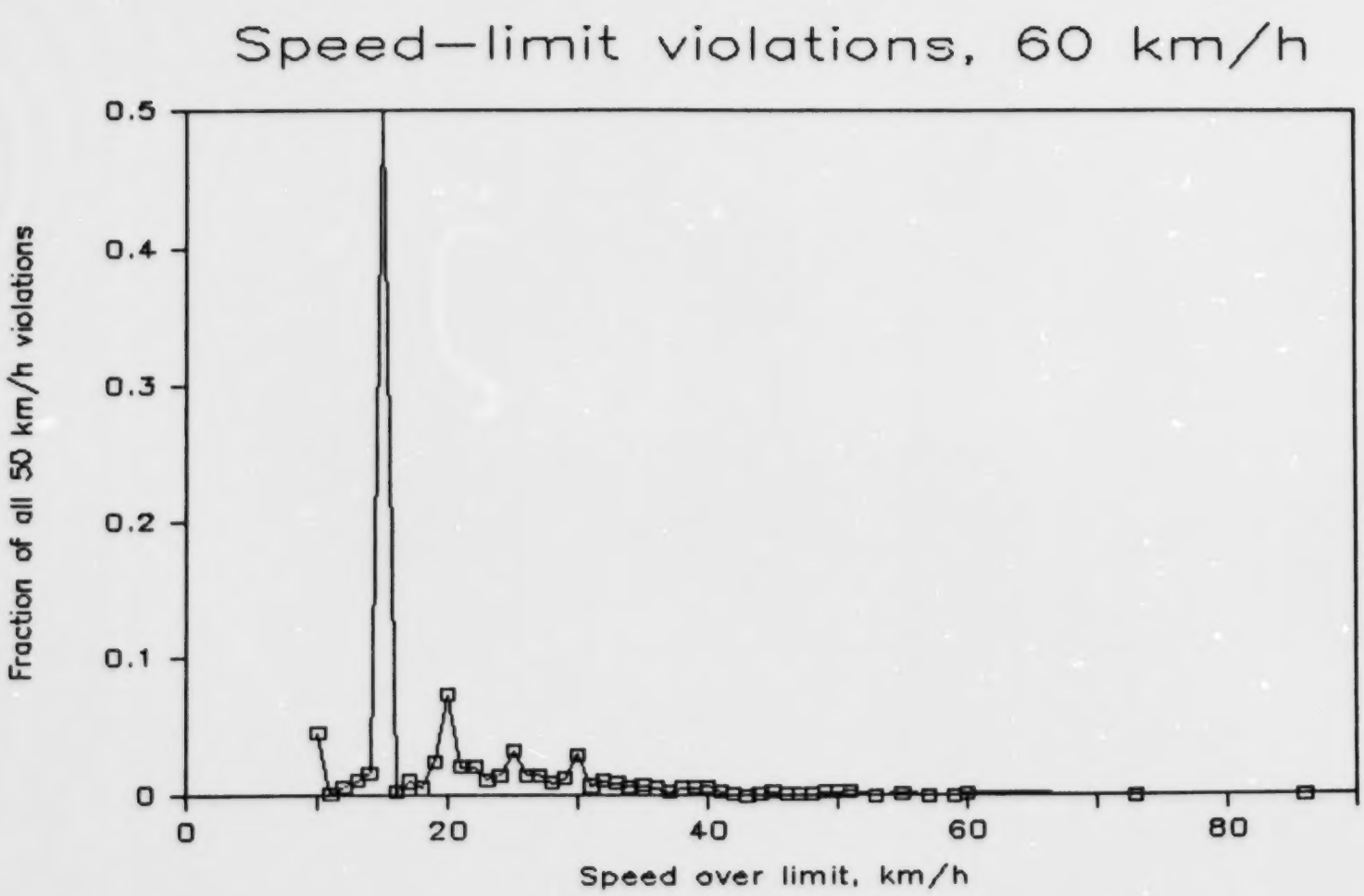

Speed—limit violations, 60 km/h
Fraction of all 50 km/h violations
0.5
0.4
0.3
0.2
0.1
0
0
20
40
60
80
Speed over limit, km/h

FIGURE 4

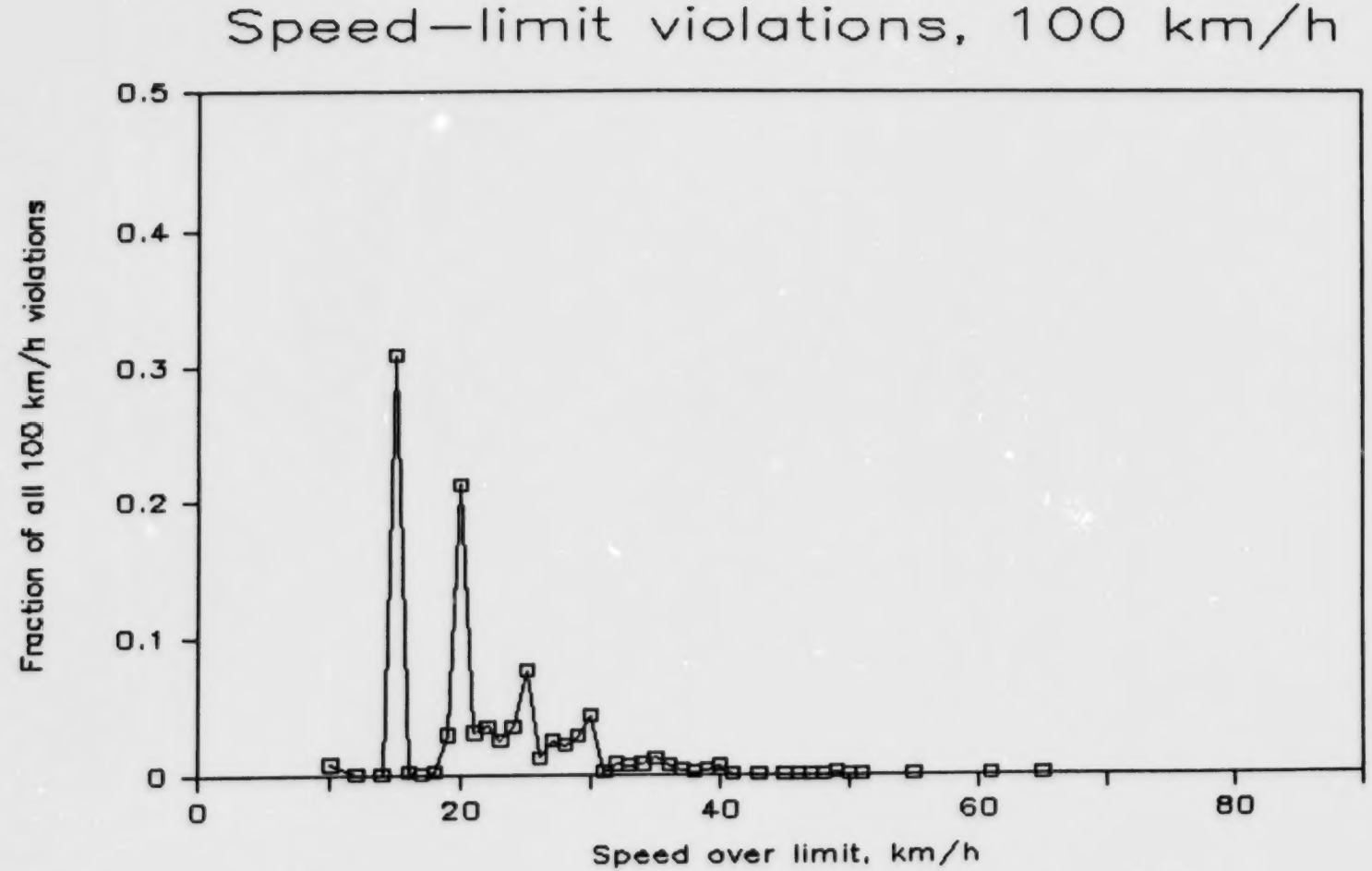

Speed—limit violations, 100 km/h
Fraction of all 100 km/h violations
0.5
0.4
0.3
0.2
0.1
0
0
20
40
60
80
Speed over limit, km/h

TABLE 6 : Accidents for drivers with 1 speeding conviction in 1 year

Speed Limit	Above Limit	No. of Drivers	3 yr. Accs.	Weighted Mean	95% Limits Upper	Lower
All	0-15	123224	33755	0.293	0.296	0.291
All	16-29	79589	22479	0.286	0.289	0.283
All	30-49	23891	7331	0.297	0.302	0.291
All	>49	894	371	0.315	0.344	0.286
20-40	0-15	24264	6307	0.288	0.293	0.283
50-70	0-15	80248	22037	0.294	0.297	0.291
80-100	0-15	18712	5411	0.292	0.298	0.286
20-40	16-29	5276	1413	0.281	0.293	0.270
50-70	16-29	39995	11427	0.293	0.297	0.288
80-100	16-29	34318	9639	0.281	0.285	0.276
20-40	30-49	683	185	0.269	0.302	0.239
50-70	30-49	9304	2914	0.305	0.314	0.296
80-100	30-49	13904	4232	0.292	0.300	0.285
20-40	>49	14	8	0.186	0.432	0.064
50-70	>49	405	183	0.341	0.386	0.299
80-100	>49	475	180	0.272	0.312	0.236

Determining Conviction Categories:Step 2 On the basis of such comparisons, it was possible to combine some of the 45 conviction categories which were found to be associated with similar average numbers of accidents. The resulting 14 conviction groupings, the associated three-year average number of accidents and 95% confidence limits of the estimate are shown in Table 7 and plotted in Figure 5. Also shown are estimates of accident potential for conviction-free drivers, those who had no convictions of any type during one calendar year.

In summary, the final 14 conviction categories to be used in analysis were established on the basis of the following considerations:

1. Conviction types within each category were similar in nature;

2. The accident potential associated with each conviction within a category was similar in size;

3. The number of drivers having offenses in each category was considered sufficient to provide a reliable estimate of accident potential for that category.

A list showing which convictions are included in the first 14 conviction categories is given in Appendix 2.

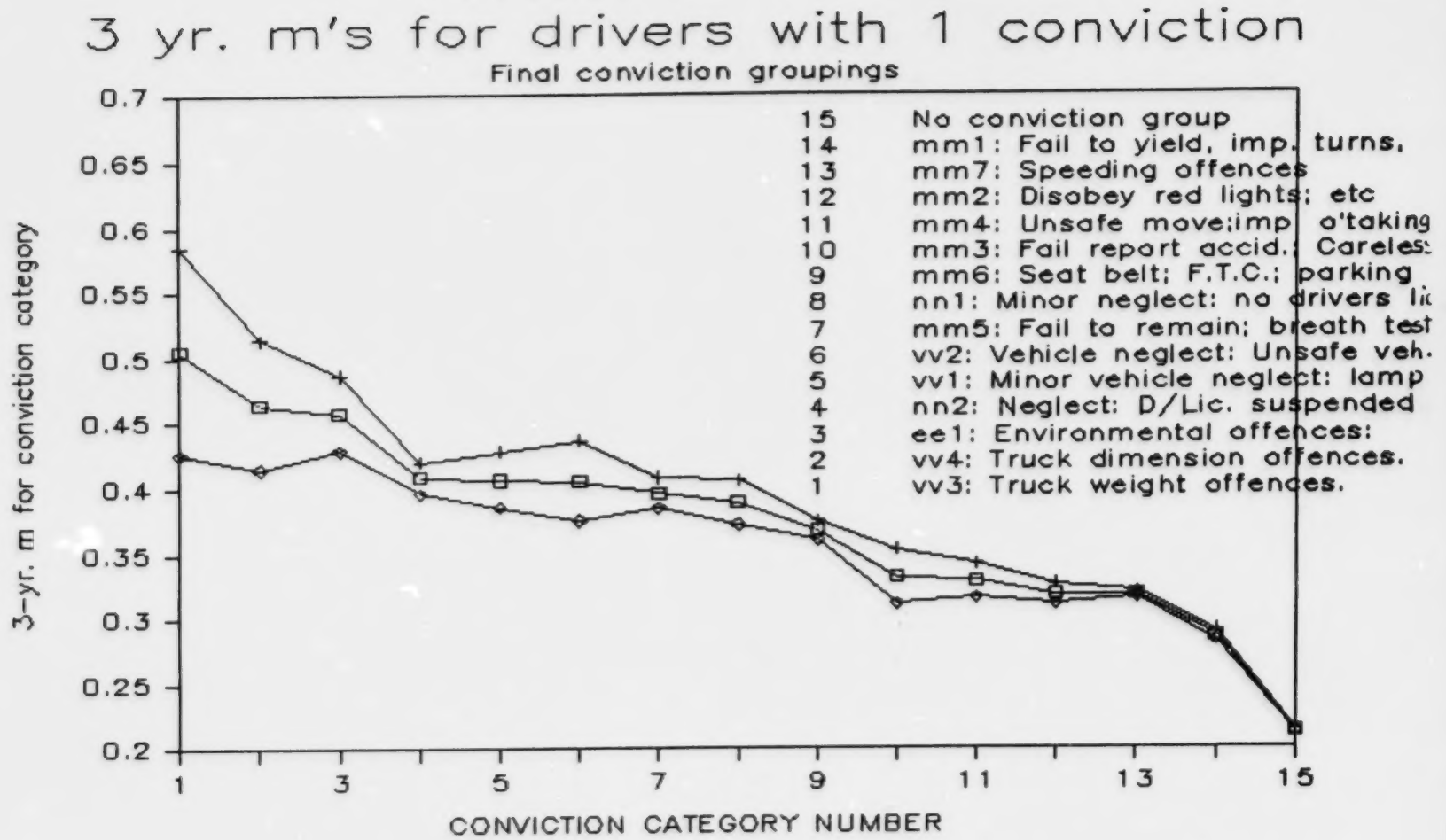
3.6 Discussion of Conviction Categories

Inspection of Figure 5 leads to the question of why it is that convictions for some of the innocuous offenses (e.g. for a burnt-out light bulb, nn1) are found to be associated with more accidents than offenses traditionally deemed very dangerous such as speeding (mm7) or running a red light (mm2). Several reasons combine to explain this apparent paradox. First, convictions associated with most accidents are those characteristic of trucks. Since trucks cover 10-20 times the distance of a passenger car per year, it is to be expected that they will have, on the average, more accidents. The second reason is easier to explain through an example. Assume that 1000 run-the-red offenses lead to 5 accidents and that 1000 fail-to-signal-turn offenses lead to 1 accident. Thus, we are assuming that to run-the-red is a more dangerous offense than failing to indicate a turn. However, the enforcement of both offenses is unequal (perhaps because one is assumed to be more dangerous than the other). Assume further that of the 1000 run-the-red offenses, 10 lead to a conviction whereas of the 1000 fail-to-signal offenses 1 leads to a conviction. Thus, a figure similar to Figure 5 would show 5/10 = 0.5 accidents per conviction for running the red and 1/1 = 1.0 accident per conviction for failure to indicate a turn. Even though, according to our starting assumption, running the red is five times more dangerous than failing to indicate a turn, because enforcement of the two offenses is unequal, the final results show the contrary. The problem is caused by the fact that the driver record contains information about the number of convictions, not the number of of illegal actions committed by a driver for which he is not convicted. This is the data with which we have to work. A third reason may be related to the connection between different types of behaviour and convictions. The incidents on a person's driving record, convictions and accidents, may be indications of their overall driving behaviour. As a result

TABLE 7 : 3-YEAR MEAN ACCIDENTS FOR DRIVERS WITH 1 CONVICTION IN 1 YEAR. (Based on Final Conviction Categories)

Category (Final)	Brief Description	Drivers	Accidents	3-Year Mean	90% Limits Upper	Lower
mm1	Fail to yield, improp.turns,etc	37775	10425	0.287	0.291	0.282
mm2	Disobey red lights etc.	13183	4169	0.319	0.326	0.311
mm3	Fail report acc; C'less driving	1774	657	0.332	0.353	0.311
mm4	Unsafe move, overtaking	4606	1514	0.329	0.342	0.316
mm5	Fail to remain; breath test	6263	2563	0.396	0.408	0.385
mm6	Seat belt; FTC; parking	15454	5885	0.367	0.375	0.360
mm7	Speeding	165852	52623	0.318	0.321	0.316
nn1	Minor neglect;plates,license	2865	1166	0.388	0.405	0.372
nn2	Neglect;d/lic. suspended	5720	2485	0.408	0.420	0.396
vv1	Minor vehicle: lamps, etc.	1802	786	0.406	0.428	0.385
vv2	Vehicle: tires, brakes, unsafe	903	441	0.405	0.436	0.375
vv3	Insecure load	134	94	0.505	0.584	0.425
vv4	Weight & dimension offences	348	214	0.464	0.513	0.415
ee1	Environmental offences	1022	527	0.457	0.486	0.428
	Drivers with no convictions	2047414	337745	0.216	0.213	0.212

FIGURE 5


3 yr. m's for drivers with 1 conviction
Final conviction groupings
3-yr. m for conviction category
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
1
3
5
7
9
11
13
15
CONVICTION CATEGORY NUMBER
15 No conviction group
14 mm1: Fail to yield, imp. turns,
13 mm7: Speeding offences
12 mm2: Disobey red lights; etc
11 mm4: Unsafe move;imp o'taking
10 mm3: Fail report accid.; Careles
9 mm6: Seat belt; F.T.C.; parking
8 nn1: Minor neglect: no drivers li
7 mm5: Fail to remain; breath test
6 vv2: Vehicle neglect: Unsafe veh.
5 vv1: Minor vehicle neglect: lamp
4 nn2: Neglect: D/Lic. suspended
3 ee1: Environmental offences:
2 vv4: Truck dimension offences.
1 vv3: Truck weight offences.

the types of convictions committed by certain types of people may also provide insight into their potential for accidents. If a person engages in certain behaviours which lead to certain convictions they may also engage in certain other behaviours which predispose them to accidents. To illustrate with an example, drivers with environmental types of convictions (e.g. noisy muffler) were found to have a higher weighted mean of accidents than drivers with most other types of convictions. Most drivers quickly have their noisy muffler fixed and are unlikely to receive this type of conviction. The attitude which results in drivers coming to the attention of the police and being charged with this offense may be related to a similar careless attitude towards behaviour which results in accidents.

All of the three reasons discussed above affect the correlation between a particular offense category and accidents. Thus, correlation does not necessarily indicate causality. Two conclusions follow. First, one should not interpret the results in Figure 5 as providing information about the danger inherent in this or that offense. Second, one should not be surprised when, later, in section 4, the same innocuous offenses will prove to be strongly related to the driver's estimated accident potential.

3.7 Age and Gender Categories

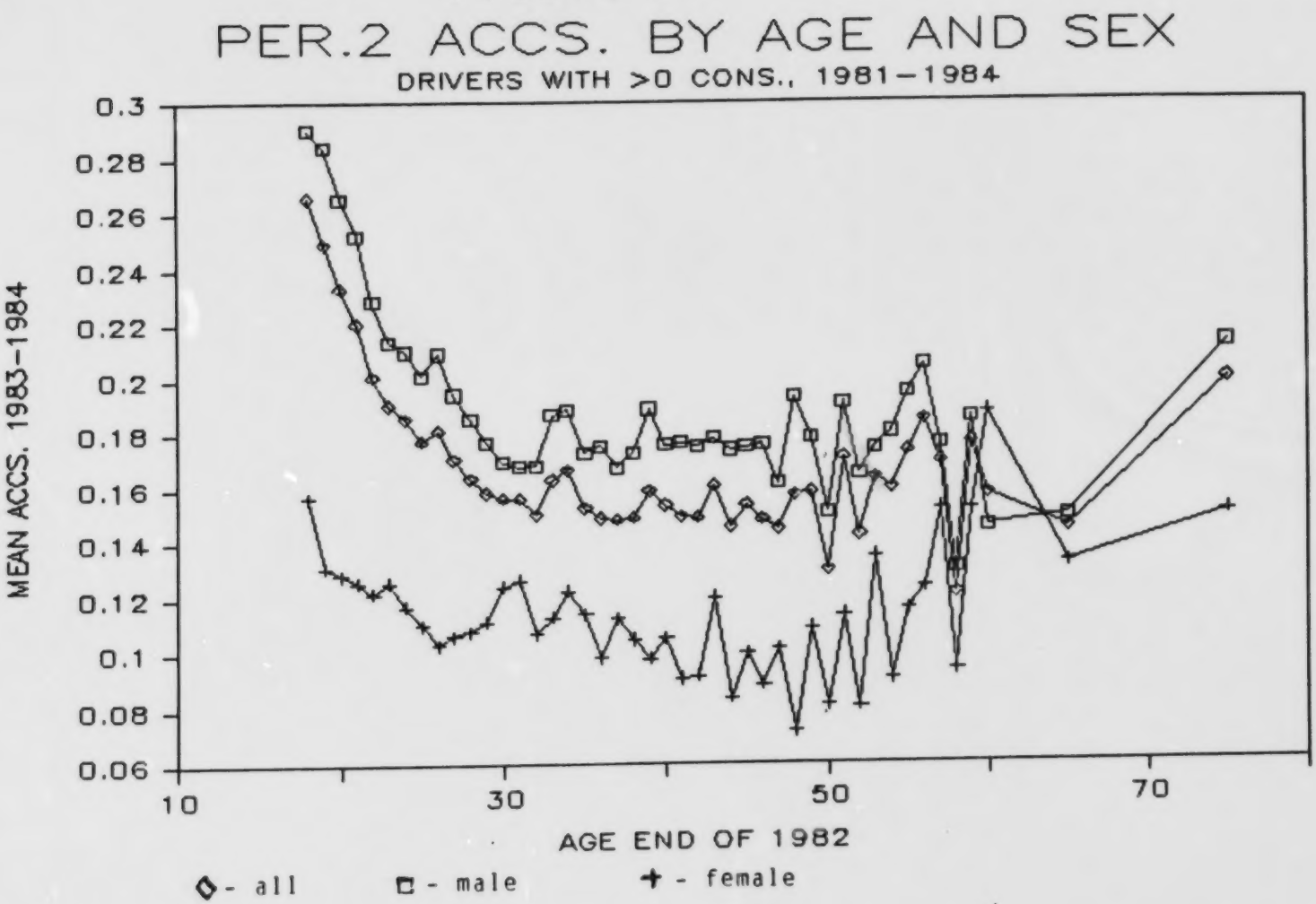
It is well known that the average number of accidents for a driver depends on gender and age. To account for this fact, age and gender will be used in the analysis as "variables". It is relatively simple to account for gender since it comes in two natural categories. However, the relationship between age and number of accidents is continuous in nature and distinctly non-linear as is shown in Figure 6. To include age in the analysis it was necessary to establish a number of age categories. After careful analysis, the boundaries between age groups were chosen so that the average numbers of accidents within each group remained nearly constant while still keeping sufficient numbers of drivers within each age category to maintain statistical reliability. The following eight age categories were chosen:

<21
21-25
26-30
31-35
36-40
41-50
51-60
>60

3.8 Exclusion of Drivers with Suspended Licenses

Some drivers in the sample had their license suspended during the study period. Most of these are drivers who had many convictions which carried demerit points. The extent to which a suspended driver curtails his or her driving is unknown. This leads to serious difficulty in the statistical analysis. Consider a driver who in the first two-year period had many convictions and was suspended. That driver can be expected to drive less in the second two-year period and therefore to have proportionately fewer accidents.

FIGURE 6


PER.2 ACCS. BY AGE AND SEX
DRIVERS WITH >0 CONS., 1981-1984
MEAN ACCS. 1983-1984
0.3
0.28
0.26
0.24
0.22
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
10
30
50
70
AGE END OF 1982
◇ - all
□ - male
+ - female

In the statistical analysis this would tend to create a negative correlation between accidents and convictions. That is, it would lead to the incorrect result that the larger the number of convictions in the first two years, the fewer accidents a person is likely to have in a subsequent period. The net effect of this difficulty is to distort the results of analysis in some unpredictable way. In fact, in the initial statistical analyses, this distortion was so large that negative weights were produced for criminal code offenses for which drivers were likely to be suspended. (The implication of a negative weight is the paradoxical result that the offense is associated with fewer subsequent accidents.) As a result of this finding, it was necessary to remove from the data set and from subsequent analysis those drivers whose license had been suspended during any time in the period 1981-1984. The result of this decision is that whatever results are obtained in the course of the analysis apply directly only to those drivers whose licenses have not been suspended. The extension to drivers who were suspended under the present demerit point system is therefore an extrapolation.

4.0 ANALYSIS AND RESULTS

The various activities described so far (checking representativeness, conducting consistency checks, the exploratory analysis of data, the selection of conviction categories, determination of age groups and the removal of suspended drivers) are all preliminary to the main activity, namely the establishment of a relationship between information contained in a driver's record and his or her expected number of future accidents. The information used was a driver's gender, age, the count of accidents (at-fault, not-at-fault or total), and the count of convictions in each of the fourteen categories. This information from the first two years was used to estimate "weights" which best fitted the accident record in the second two-year period. Weights which are associated with each conviction category can be taken to be relative numbers of "points" which should be assigned for the best prediction of the likelihood of an accident in the second two-year period. Only records of those drivers who had at least one conviction in the first two-year period were used (n ~170,000).

4.1 Schemes and Variants Examined

The MTO was interested in a number of variants, each using different sets of variables. These fall into three categories:

1) models which made use of age and gender information and models which did not;

2) models which assigned points for period 1 accidents (with the further distinction between at-fault, not-at-fault and total), and models which did not assign points for accidents; and

3) models which assigned different numbers of points for each conviction category and models in which all convictions carried the same weight.

In total 16 different combinations of variables from the first two-year period were used. For each of these, "weights" were estimated. Each of these 16 combinations results in a prediction equation which is termed a "model".

Table 8 shows which variables were used in each "model". Thus, for example, model B2 calculates the expected number of period-two accidents using the count of period-one convictions in each of the 14 types (see Table 7) as well as the count of total period-one accidents; model C4 uses driver age and gender, the count of all convictions (without distinction with respect to type), the count of at-fault accidents and the count of not-at-fault accidents.

Half of the models tested (A1-A4, B1-B4) allowed for a different weight to be given to each of 14 conviction types. The other 8 models tested used a weighting of 1 point per conviction, that is, all 14 conviction types were considered equivalent (models C1-C4, D1-D4). Half of the models used age and gender as variables, to aid in the prediction of accidents in the second two year period (models A1-A4, C1-C4).

One quarter of the models (A1, B1, C1 and D1) used only information on convictions in the first two year period to predict accidents in the second two year period. One quarter of the models (A2, B2, C2 and D2) used total accidents in the first two year period as well as convictions, to predict subsequent accidents. One quarter (A3, B3, C3 and D3) used only at-fault accidents and convictions for prediction purposes. Finally, one quarter (A4, B4, C4 and D4) used separate weights for at-fault and not-at-fault accidents, as well as convictions (also separately weighted: A4, B4 or one point per conviction: C4, D4) to predict subsequent accidents.

4.2 What is a "Model" and how does it "Work"

By "model" we mean a mathematical equation which allows the estimation of a driver's expected number of accidents in "period two" on the basis of that driver's age, gender, record of convictions and record of accidents in "period one". Like many others, we have elected to use the linear-sum form.

To illustrate, consider a female driver, 24 years of age who in the first period had two speeding convictions, one conviction for failing to yield the right of way (ROW), and one at-fault accident. To estimate her expected number of period two accidents we consider the importance of each variable.

Variable	Weight	Count	Contribution
Base Driver	0.176	-	+0.176
For being female	-0.061	-	-0.061
For being 24 yrs.	-0.039	-	-0.039
For speedg. conv.	0.027	x2	+0.054
For ROW convic.	0.027	x1	+0.027
For accident	0.058	x1	+0.058
Estimate of expected number of period two accidents			=0.215

TABLE 8: VARIABLES USED FOR REGRESSION RUNS
(x -- indicates variable used in run)

Run	Age and Sex dummy variables	Variables for each Conviction Group	Variable for total Convictions	Accident variables		
				Total	Fault	At Fault
A1	x	x				
A2	x	x		x		
A3	x	x			x	
A4	x	x			x	x
B1		x				
B2		x		x		
B3		x			x	
B4		x			x	x
C1	x		x			
C2	x		x	x		
C3	x		x		x	
C4	x		x		x	x
D1			x			
D2			x	x		
D3			x		x	
D4			x		x	x

The "weights" in this illustrative example are for Model A2. (See Table 9.) Thus, we estimate that females who were 24 during 1981 and had during that period of time the aforementioned record of convictions and accidents, should be expected to have 0.215 accidents during 1983-1984 on average.

4.3 On the Estimation of "Weights" and "Diversity"

The main statistical work is the estimation of "weights". These were estimated for each model separately. For example, the weights for model A2 are given in the "Estimated Coefficient Column" of Table 9. The magnitude and relative importance of the weights for model A2 is shown in Figure 7. A full collection of such tables for all 16 models is provided in Appendix 3. In Appendix 4 we give fuller technical details about the distinctive features of the estimation procedure and the software used.

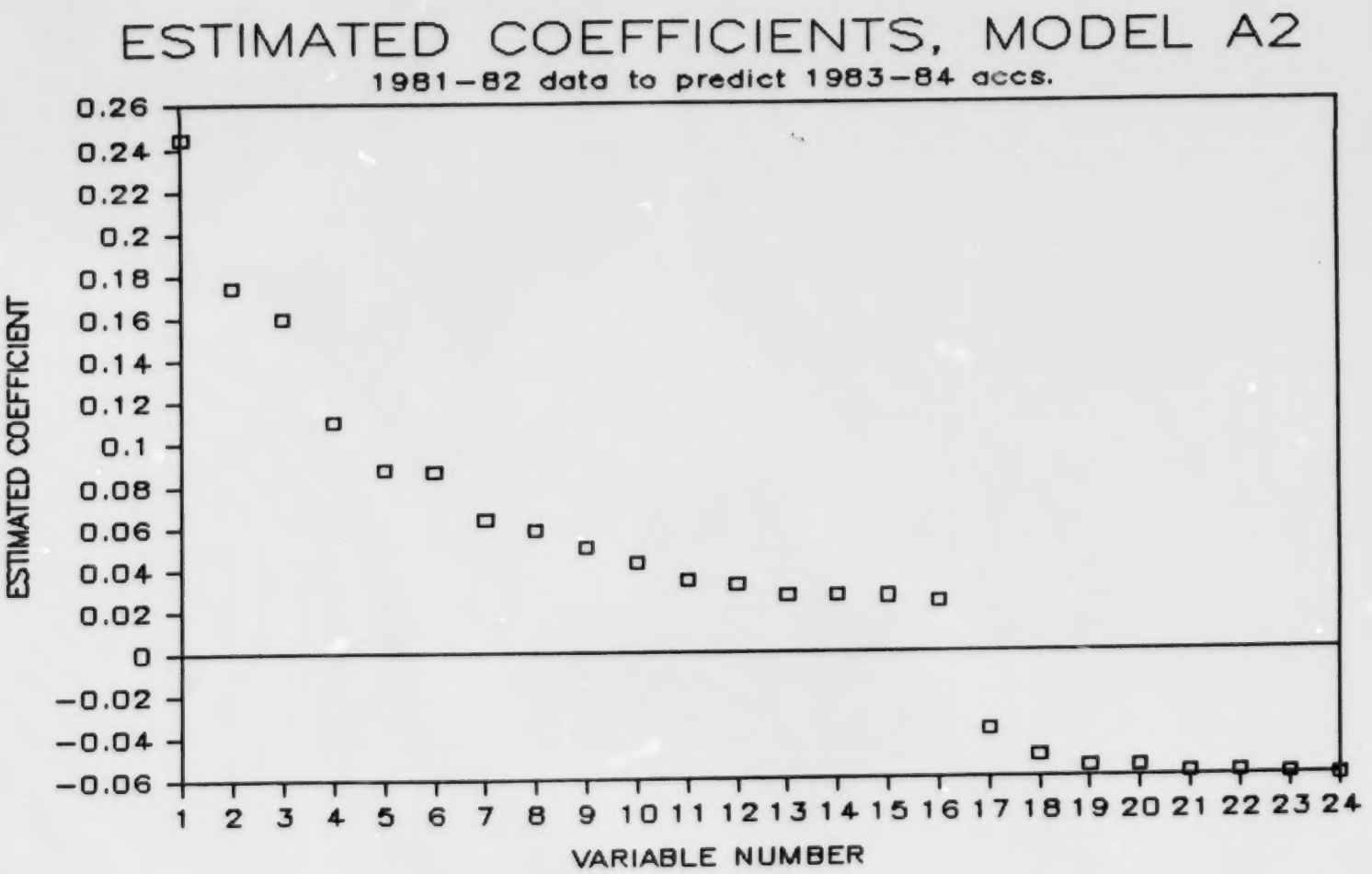
The model weights allow us to estimate the expected number of accidents of an average driver with known age, gender and record of convictions and accidents. However, while this is true for an "average" person fitting such traits, any specific driver in this group will have her own "expected number of accidents". Our statistical procedure also allows us to describe the diversity within each group. Having obtained estimates of the "weights" for a model we have examined the data again in order to establish the standard deviation of what we estimate to be the expected number of accidents for a driver with given traits and record. It turns out that the estimate of standard deviation is equal to the estimate of expected number of accidents multiplied by a constant. The list of "constants" for the 16 models is given below.

Model	Constant
A1	0.67
A2	0.64
A3	0.66
A4	0.64
B1	0.71
B2	0.68
B3	0.69
B4	0.68
C1	0.67
C2	0.65
C3	0.67
C4	0.65
D1	0.72
D2	0.69
D3	0.71
D4	0.68

TABLE 9: Regression coefficients for Model A2

Description of variable	Estimated Coefficient	Standard Error
Intercept (Male, Age < 21)	0.1763	0.002876
Dummy variable for age 21-25	-0.03856	0.002852
Dummy variable for age 26-30	-0.0565	0.002913
Dummy variable for age 31-35	-0.06006	0.003044
Dummy variable for age 36-40	-0.05202	0.003516
Dummy variable for age 41-50	-0.06013	0.00351
Dummy variable for age 51-60	-0.05774	0.004832
Dummy variable for age > 60	-0.06156	0.007068
Dummy variable for female	-0.06122	0.001735
mm1: Fail to yield, imp. turns, PXO, amber violations, etc	0.02696	0.001753
mm2: Disobey red lights; rail crossing violations.	0.042125	0.002916
mm3: Fail report accid.; Careless driving; Dang. driving;Crim. neg. caus. death	0.023304	0.008613
mm4: Unsafe move;imp. o'taking; Disobey signs	0.06359	0.004959
mm5: Fail to remain; breath test; alcohol; impairment.	0.2444	0.054575
mm6: Seat belt; F.T.C.; parking; divided h'way offences.	0.03221	0.00213
nn1: Minor neglect: no drivers license; permits; insurance, address change	0.02624	0.010463
nn2: Neglect: D/Lic. suspended or not produced; plates, insurance.	0.03366	0.003708
vv1: Minor vehicle neglect: lamps windows obstructed, etc.	0.04954	0.006561
vv2: Vehicle neglect: Unsafe veh., brakes, tires	0.08673	0.009296
vv3: Insecure load	0.159215	0.02721
vv4: Weights and dimension offences	0.11074	0.011895
ee1: Environmental offences: noise, fumes.	0.08748	0.008334
mm7: Speeding offences	0.026515	0.000866
Total accidents in period 1	0.05831	0.001580

FIGURE 7


ESTIMATED COEFFICIENTS, MODEL A2
1981–82 data to predict 1983–84 accs.
ESTIMATED COEFFICIENT
0.26
0.24
0.22
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
−0.02
−0.04
−0.06
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
VARIABLE NUMBER

Thus, for example, the standard deviation of "expected number of accidents" for females of the aforementioned group is estimated to be 0.215x0.64 = 0.138. The "constants" above are an index of model performance. The larger the constant, the larger the standard deviation, and the less accurately we know the accident potential of a certain driver. The ability to describe the diversity of "accident potential" within specific groups of drivers will prove important when we analyze the performance of the current demerit point scheme and compare it to the performance of other "models".

4.4 The Accident Potential of Ontario Drivers

The models estimate, for each driver, the "number of accidents he or she is expected to have per year". This is not an integer because its nature is that of a "long-term-average". For brevity this number is called a driver's accident potential. Of course, not all drivers have the same accident potential; some drive more, some drive less, some take risks, others are more cautious. Prior to examining the results for each model, let us examine the diversity of accident potential in the population of Ontario drivers. This will reveal how many drivers there are in the population who have a high accident potential. How many of these "high accident potential" drivers will indeed be identified for post-licensing-control under the current demerit point system and the new weighting models, will be examined later.

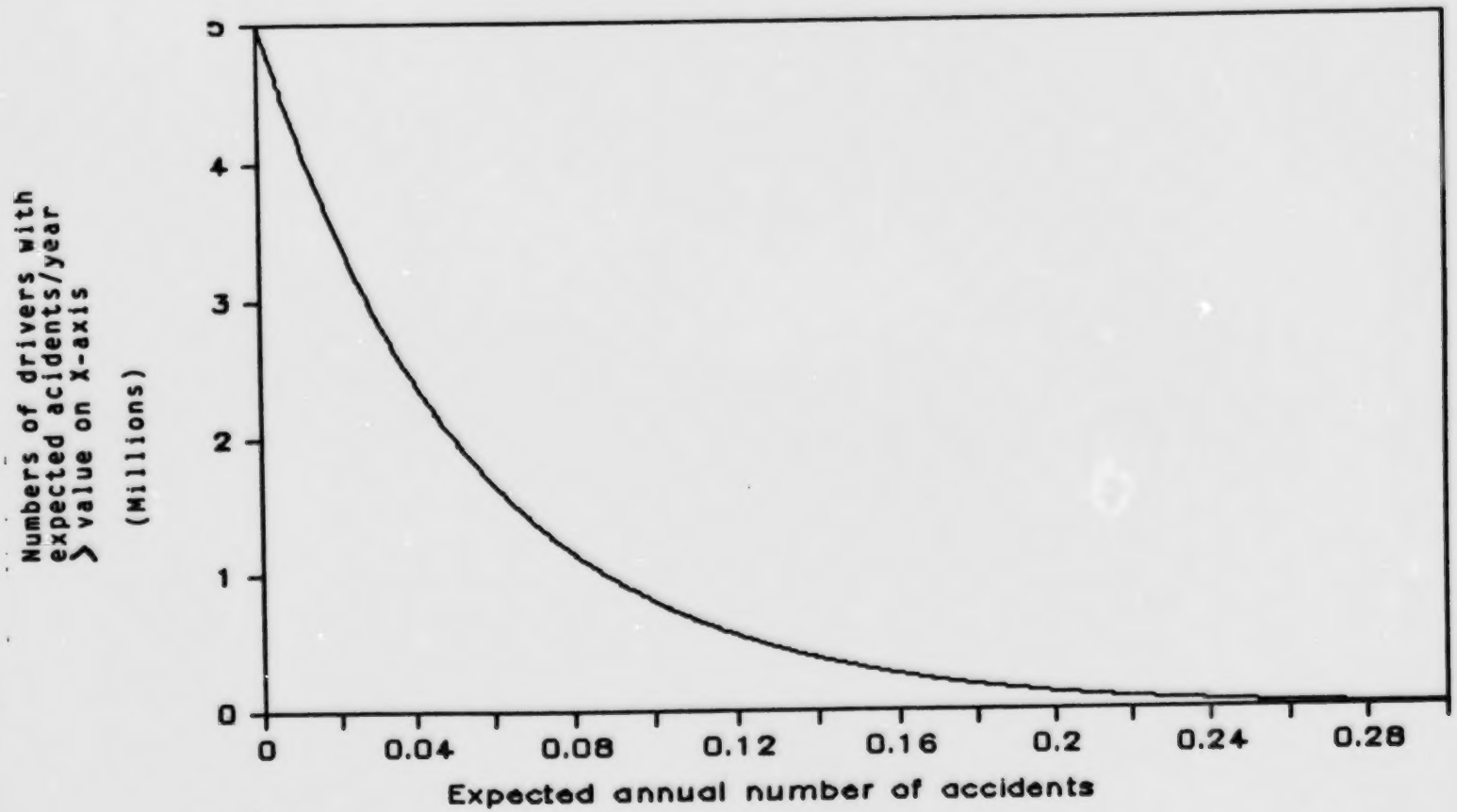
The number of accidents in the second two-year period was used to compute the mean accident potential (0.055 accidents/year) and standard deviation of accident potential (0.055 accidents/year) in the total driver record sample. Details of the method are given in Appendix 4.3. This information was then used to plot the distribution of accident potential in a population of 5 million Ontario drivers shown in Figure 8.

Using Figure 8 one can estimate how many drivers in the population are at each level of accident potential. Thus, for example, it can be seen that about 1.75 million drivers, or 1/3 of the drivers, have an accident potential which is higher than the average. (It is less than 1/2 of the population because there are some drivers with a very high accident potential). Let us now find out how many drivers have an accident potential which is much higher than the mean for this population. Consider drivers whose accident potential is 3 standard deviations above the mean or higher, that is, an accident potential larger than 0.22 accidents/year. From Figure 8 it can be seen that almost 90,000 drivers have an accident potential of 0.22 or higher.

It is drivers with a high accident potential who are to be identified for post-licensing-control. However, it must be noted that a driver may have a high accident potential because of bad driving as well as because of high exposure to risk. If a driver is likely to have x accidents driving y kilometers, the same driver is likely to have 2x accidents driving 2y kilometers. Thus, it is very important to ascertain whether one has identified, through the demerit point system, a driver who drives little but has a high accident potential or one who drives a lot and has a correspondingly high accident potential. The two drivers may require different treatments. For brevity we will denote accident potential by "m".

Figure 8: Distribution of accident potential in the Ontario Driving Population

mean = 0.055 (±0.055) accidents/year



Ordinarily one is interested in the identification of drivers with unusually high m's. Figures 9-11 are successive enlargements of the right-hand tail of the curve in Figure 8. The reader must keep in mind the fact that the shape of the curve reflects the assumption that the distribution of m's is Gamma. We cannot be sure that this assumption is valid until we show that we can reproduce the actual accident data by using it. For example, if accidents were distributed in a bimodal fashion because people drove either a little or a great deal, and had either very few or a lot of accidents, then our assumption of a Gamma distribution would be incorrect and we would not be able to reproduce what is observed in terms of accident distribution by using the assumption. However, if accident data is distributed in a unimodal fashion then our experience is that the gamma distribution is a good approximation.

In Appendix 4 (Tables A4.1 through A4.25), we show the results of the check on the validity of our assumption of a Gamma distribution. For example males aged 21-25 years (Table 4.12) have a mean number of accidents in period one of 0.2049, and a variance of 0.2231. Using the Gamma distribution, with this mean and variance results in predictions which closely match what has been observed. Thus, for example, for drivers with 0 accidents in period one, we predict that there will be 45249.8 drivers who have 0 accidents in period two. The observed number of such drivers was 45054. The similarity between observation and prediction is consistent for the various age and sex groups This indicates that the Gamma assumption is reasonable.

The information contained in Figures 8-11 enables one to form an idea about the number of drivers with unusually high m's in the population. The problem is, of course, that these drivers are difficult to identify on the basis of their conviction and accident record. A driver may have a relatively low m while accumulating a large number of convictions and even accidents in a certain period of time. Conversely, a person who, in a very long period of time, would be recognized as having an unusually high m, might in a one or two year period have no accidents and no convictions. The models obtained in Section 4 can be thought of as a net; their purpose is to fish out those drivers who have an unusually large m. However, even the best models formulated in Section 4 are but an imperfect net for the identification of high-m drivers. Many such drivers will slip through its holes; many of those identified will not be high-m drivers. The degree of success and failure of the net will be examined next.

4.5 The performance of the current demerit point system and of the new models

Since the 16 new models were derived using appropriate statistical methods, rather than weighting each offense according to its perceived seriousness, they should perform better than the current demerit point system. However, all models face the same inherent difficulties as the current demerit point system. Namely, because of the randomness inherent in the process of accident occurrence and the randomness inherent in the process by which drivers acquire convictions, a two year record is just too short to accurately estimate a driver's accident potential. As will be seen, the new models are an improvement on the current system, but like the current system, still fail to detect most of the high accident potential drivers.

FIGURE 9

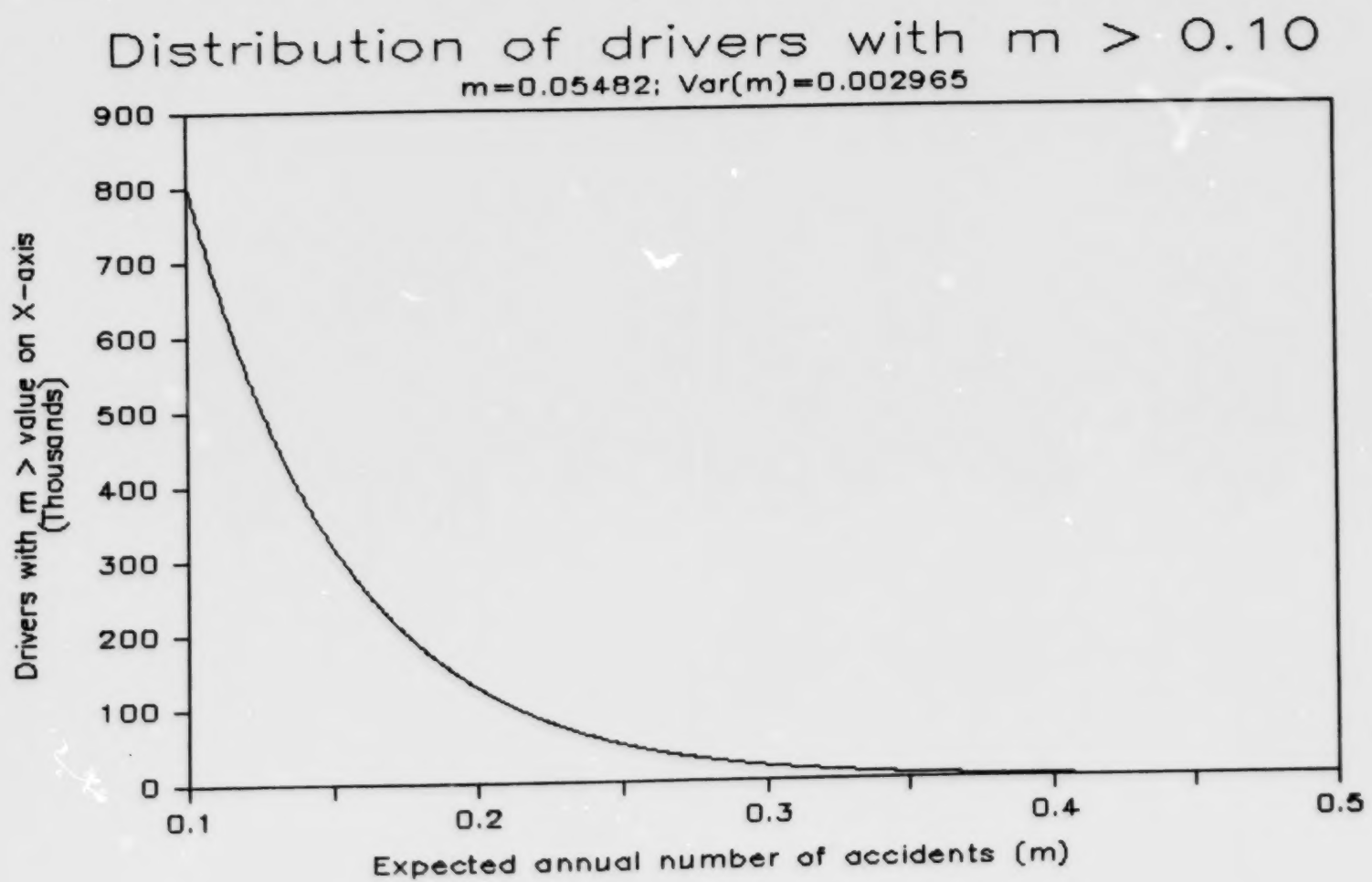

Distribution of drivers with m > 0.10
m=0.05482; Var(m)=0.002965
Drivers with m > value on X-axis (Thousands)
900
800
700
600
500
400
300
200
100
0
0.1
0.2
0.3
0.4
0.5
Expected annual number of accidents (m)

FIGURE 10

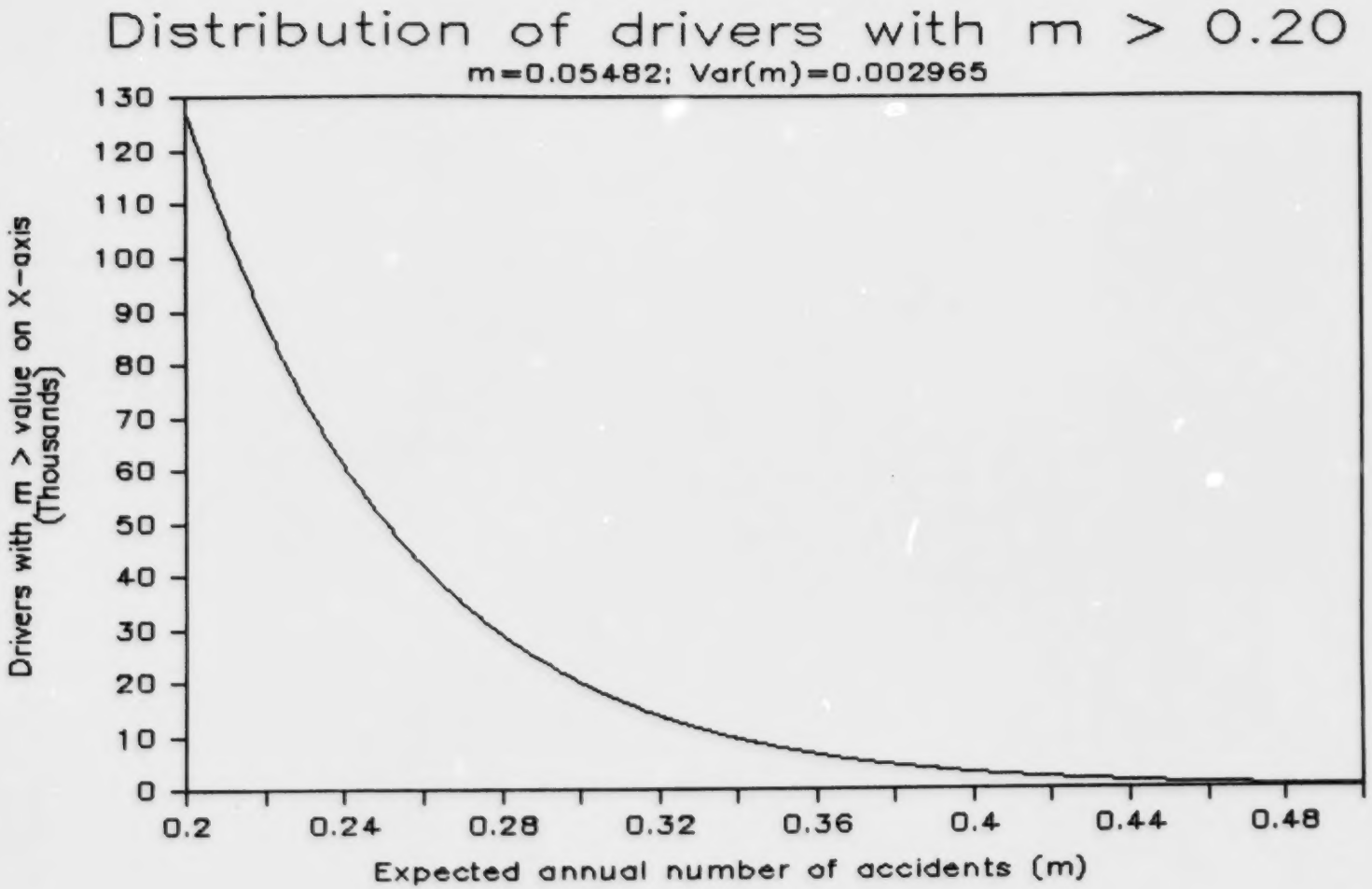
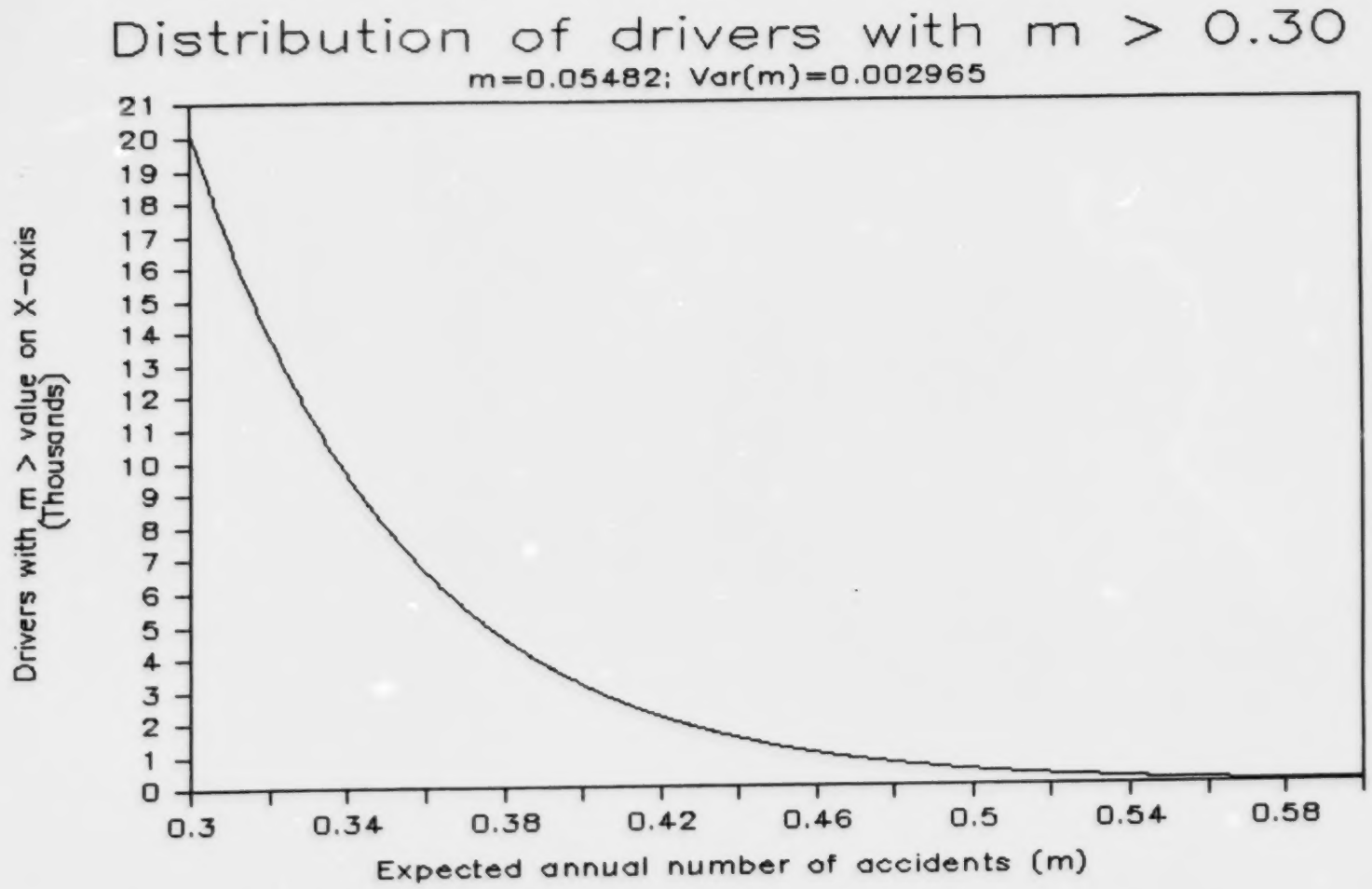

Distribution of drivers with m > 0.20
m=0.05482; Var(m)=0.002965
Drivers with m > value on X-axis (Thousands)
130
120
110
100
90
80
70
60
50
40
30
20
10
0
0.2
0.24
0.28
0.32
0.36
0.4
0.44
0.48
Expected annual number of accidents (m)

FIGURE 11


Distribution of drivers with m > 0.30
m=0.05482; Var(m)=0.002965
Drivers with m > value on X-axis (Thousands)
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
0.3
0.34
0.38
0.42
0.46
0.5
0.54
0.58
Expected annual number of accidents (m)

Two measures of performance will be used to judge the quality of a model. The first is the number of period two accidents which have occurred to 10,000 drivers selected as worst by a model. The model which identifies drivers with most period two accidents will be judged the best.

The second measure of performance gives more detailed information. It tells us how many of the drivers estimated to have a high accident potential on the basis of their period one record are truly high accident potential drivers. These are the "hits". We also learn how many of these drivers have a true accident potential which is lower than the population average. These are the "false alarms".

The first measure of performance is straightforward. Consider, for example, the 10,000 drivers who in the first two year period had most demerit points. Checking the accident records of the same drivers, we find that during the second two year period they had 1452 accidents per year (see sum of entries in last row of Table 10).

Consider now another group of 10,000; this time those who who in the first two year period had the most accidents. For this second group we find that they recorded in the subsequent two year period 1828 accidents per year (1828 is the sum of the first three entries for accidents, that is, 312 in the first 1000, 756 in the next 4000, and 760 in the next 5000). Evidently, it is better to identify drivers by their previous accident record than by previous demerit points. Imagine now that a third group of 10,000 drivers is identified, this time those for whom model A4 estimates the highest accident potential on the basis of their age, gender, as well as convictions and accidents in the first two year period. This group has 2084 accidents per year in the second period. Thus, selection by model A4 gives a richer catch than either selection by previous accidents or by the current demerit points. In interpreting these results one has to keep in mind that the count of accidents is always subject to random fluctuations.

Several conclusions emerge. First, one can do better than to use the current demerit point weights. Second, it is important to make use of the driver's record of accidents. This is already evident from the comparisons of the last two rows. However, it also emerges from the poor performance of models A1, B1, C1 and D1 which do not make use of accident data. In fact, the top 1000 drivers can be well identified by their previous accident record alone. Third, the more drivers that are identified, the lesser the "yield". Thus, the top 1000 drivers have an accident rate of ~0.3 accidents/year which is approximately 6 times the population average; for the first 10,000 drivers the average accident rate is ~0.2 and so on.

The first measure of performance, examined above, leaves the impression that the drivers which are identified indeed have an accident potential which is substantially larger than that for the population of all drivers. While this is true for the group "on the average", this group itself may not be a homogenous one. The second measure of performance by which the quality of the alternative models is to be judged relates to the diversity of accident potential within the group of drivers which these models identify.

Table 10

ACCIDENTS PER YEAR RECORDED BY DRIVERS SELECTED BY VARIOUS MODELS

Model	Drivers estimated by model to be in					
	Top 1,000	Next 4,000	Next 5,000	Next 10,000	Next 100,000	Total
A1	188	712	904	1660	13308	16772
A2	324	856	936	1798	14192	18016
A3	276	736	860	1712	13700	17284
A4	320	868	896	1736	14060	17880
B1	212	704	824	1548	12888	16176
B2	320	832	980	1628	13688	17448
B3	272	748	804	1712	13336	16872
B4	304	876	956	1636	13684	17456
C1	208	744	760	1592	13016	16320
C2	356	808	900	1672	13732	17468
C3	276	780	780	1424	13684	16944
C4	356	804	928	1616	13748	17452
D1	176	748	688	1432	12260	15304
D2	352	824	784	1608	13268	16836
D3	244	756	852	1360	13084	16296
D4	364	840	788	1576	13216	16784
Accs.*	312	756	760	1432	10780	14040
DP**	180	640	632	Not	Available	

* - Drivers with the highest accident counts in period 1 were selected
** - Drivers with the highest demerit points acquired in period 1

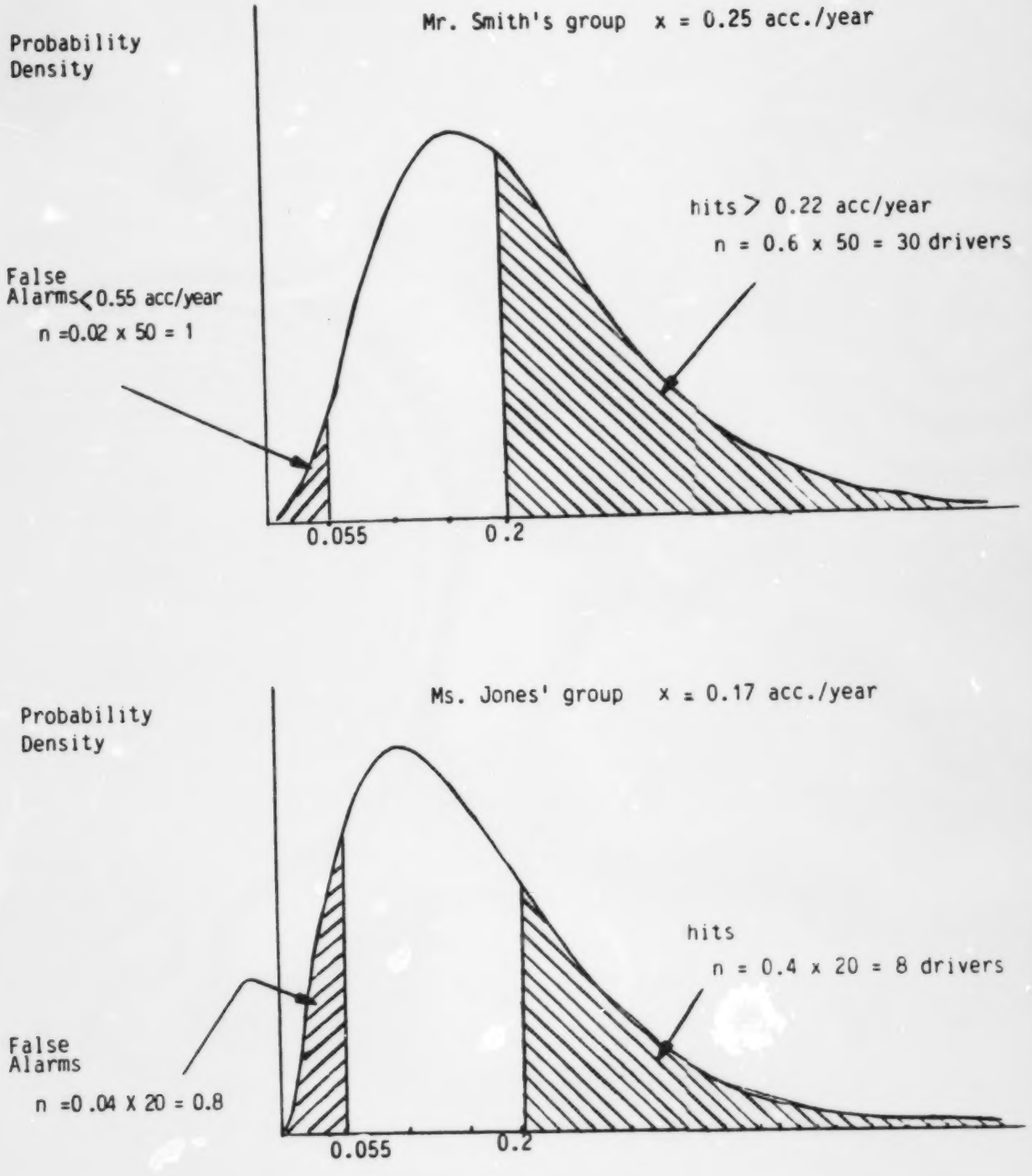
Second Performance Measure: Identifying Hits and False Alarms: A weighting scheme is like a net with which we attempt to catch drivers who, based on their two year record, are likely to have an unusually high number of accidents in the next two years. For the illustration here, consider "unusually high" to be 3 standard deviations above the mean. That is, we are hoping to identify drivers whose accident potential is larger than 0.22 accidents per year. If we manage to identify such a driver we will call this a "hit". Conversely, if based on the two year record we identify, and call in for treatment, a driver whose accident potential is lower than what is the average accident potential in the population, that is 0.05 accidents per year, we will call this a "false alarm".

Determining the hits and false alarms for each weighting scheme, or model, is a lengthy and rather complex procedure. It is not simply a matter of counting how many drivers had accidents in the second two year period, and how many did not. An explanation of the process, in general terms, follows. Let us consider one driver, out of a sample of 827,995, Mr. Smith. Mr. Smith's record of convictions and accidents in the first two year period is used to estimate his expected number of accidents per year in a subsequent time period, using one of the point schemes or models that have been developed, for example, A4. (See section 4.2 for details about the estimation of expected numbers of accidents using a driver's record and model A2). This process is then repeated for all other drivers. Now let us select those 1000 drivers who are estimated (say, using model A4), to have the highest expected number of accidents. (In the current demerit point system this is equivalent to selecting the 1000 drivers who have the highest number of demerit points). Let us suppose that Mr. Smith is among these.

Among the 1000 drivers there will be other drivers of the same age group, sex, and record as Mr. Smith. These drivers, though similar to Mr. Smith in these three characteristics, will drive different amounts, in different driving styles and are expected to have different numbers of accidents. For example, some who are in fact good drivers may have been in accidents or received convictions as a result of acts they seldom commit. These are expected to have few accidents in a subsequent period. Others in this group are in fact poor drivers and are expected to have a large number of accidents in the subsequent period. Thus the population of drivers who are in the same age group, sex and driving record as Mr. Smith in the first two year period, in fact have very different expected numbers of accidents. However, to the model, these people appear to be the same and thus are assigned the same *estimated expected number of accidents.*

The model reflects the diversity of real accident potential in this group, by not only telling us the *average* expected number of accidents, but also the *variance* of expected number of accidents in this group. Thus for Mr. Smith and drivers of the same age group, sex and record there is a distribution of expected number of accidents which is estimated by our model. As shown in Figure 12 for this group of drivers, with an average expected number of accidents of 0.25, 60% in reality have an expected number of accidents of 0.22 (hits) or more while 2% at the other end of the distribution have in reality an expected number of accidents of 0.055 or less (false alarms).

Figure 12



Note: These graphs are for the purpose of illustration only and are not meant to accurately represent particular data sets.

Now consider another driver, Ms. Jones, who belongs to a different group of drivers amongst our population of 1000 drivers with the worst records. In this group the average expected number of accidents is different from Mr. Smith's group (though it is still high), and the curve of the distribution of actual expected numbers of accidents is different (See Figure 12). Here it may be that 4% of the drivers have a low expected number of accidents of 0.055 or less (we call these *false alarms*), while 40% have an expected number of accidents of 0.22 or more (which we call *hits*).

In order to determine the total number of hits (drivers with an expected number of accidents of 0.22 or more) in the population of 1000 drivers with the worst records, we add the numbers of drivers in each of the various groups (the Mr. Smith group, the Ms. Jones group etc.) who would have more than 0.22 expected accidents. Let us assume there are 20 drivers in Ms. Jones group, and 50 drivers in Mr. Smith's group. In the case of Mr. Smith group there will be 60% of 20 drivers, or 12 hits. In the case of the Ms. Jones group there will be 40% of 20 drivers, or 8 hits. By adding up the hits we obtain the numbers in Table 11, that is 528 hits in the 1000 worst records selected by model A4, 1568 hits in the next 4000 worst records etc. Similarly, we can add up the false alarms in each of our distributions.

For illustration purposes, models A4 and B4 were selected. (The variables which these use to estimate accident potential are shown in Table 8.) Drivers were ranked in terms of accident potential estimated by each model, based on their record during the first two-year period. Table 11 shows the hits (drivers correctly determined to have accident potential larger than 0.22 accidents/year) for the 10,000 drivers from a population of 5 million Ontario drivers with the highest accident potential as estimated by each model.

As Table 11 shows, in a group of the worst 1000 drivers as selected by model A4, there are 528 hits and 39 false alarms. When drivers are selected by model B4, the number of hits increases to 541 but the number of false alarms correspondingly increases to 45. Thus, there is not much to differentiate the two models for the top 1000 drivers, probably because they consist largely of the same people. In the second 10,000 worst driver records, A4 gives 2679 hits and 933 false alarms while B4 gives 2657 hits and 1110 false alarms. Thus, A4 is not much better than B4 in terms of identifying the hits but is substantially better in terms of calling in fewer drivers which are better than average; A4 tends to have fewer false alarms than B4.

Note that the yield of hits is above 50% in the first 1000 drivers and only ~15% in drivers ranking between 20,000 and 120,000. Similarly, there are only 3-4% false alarms in the top 1000 drivers and this increases to 14-17% in the last row. Thus the larger the group identified, the worse the performance of the net.

Now let us consider another aspect of how our net performs. How many of the drivers in the population who have an unusually high accident potential will slip through the holes? Earlier we showed that there are drivers with bad records who belong to a group with a high average accident potential, but who themselves belong to the end of the distribution where the probability of an accident is low. Similarly there are many drivers with average or good driving records who belong to groups with low average accident potential but

TABLE 11 : Figures of merit for Models A4, B4

Drivers estimated by model to be in		Number of drivers expected to have m > 0.22		m < 0.05	
		Model A4	Model B4	Model A4	Model B4
the top	1,000	528	541	39	45
the next	4,000	1568	1595	246	289
the next	5,000	1601	1620	390	458
the next	10,000	2679	2657	933	1110
the next	100,000	15987	15291	14198	16928
TOTALS	120,000	22363	21704	15806	18830

who themselves are in the upper end of the distribution and have a high accident potential. Such individuals will slip through the net.

Looking again at Figure 8, it can be seen that out of 5 million drivers, 90,000 drivers have an accident potential larger than 0.22 accidents/year. Thus, as Table 11 shows, using model A4 to select the 10,000 drivers with the worst records will catch 3697 of the high accident potential drivers; calling in the next 10,000 will identify 2679 more hits. Calling in the next 100,000 will yield another 15,987 hits. Thus even after those 120,000 drivers of 5 million who have the highest accident potential according to model A4 have been selected for treatment, only 22,363 hits can be expected. Of the 90,000 drivers in the population who have an unusually high accident potential (>0.22), still 67,637 remain unidentified. One would have to call in the whole driver population before all the hits would be identified.

Table 12 compares performance among the 16 models, and the current demerit point system, in terms of hits and false alarms for the worst 10,000 drivers identified by each model. It should be noted that, while there will always be considerable overlap between groups of drivers identified by different models, there also will be systematic differences. Thus, for example, the *use of model series A and B will lead to groups which contain more truck drivers than the current system*, simply because the current system does not assign any points for truck weight or truck dimension offenses, while models A and B weigh these heavily.

While comparison in terms of hits and false alarms is good for purposes of illustration, it depends on a rather arbitrary definition of what is to be considered an "unusually high" accident potential. A more comprehensive way to characterize the performance of different models is by continuous curves, as shown in Figure 13.

In Figure 13, accident potential is measured on the horizontal axis. The lowest curve represents the current demerit point scheme. In a group of 10,000 drivers who in a population of 5 million have the most demerit points, one can expect to find 2800 who have an accident potential above 0.2 accidents per year. The highest curves represent models A4 and B4. In a group of 10,000 drivers who in a population of 5 million have the highest estimated accident potential by model A4, one can expect to find some 4200 drivers who have an accident potential above 0.2 accidents per year. Thus, the higher the curve, the better the net.

While the current demerit point scheme is clearly inferior to the new schemes, there is not much to distinguish between the other four models in Figure 13. A4 and B4 give a separate weight to each of 14 convictions and are somewhat better than C4 and D4 which assign the same weight to all convictions. A4 (which uses age and gender information) is slightly better than B4 where accident potential is less than 0.15 accidents/year. B4 is slightly better than A4 for accident potential greater than 0.15 accidents/year.

In Figure 14 we show a second group of 10,000 drivers, those who rank in the 20,000 to 30,000 worst records. That is, these drivers' records are bad, but not as bad as the 10,000 shown in Figure 13. The difference in the performance of the four models appears to be even less important.

TABLE 12 : Figures of merit for 10,000 drivers with highest m's (for each model)

Model	Number of drivers expected to have m > 0.22	m < 0.05
A1	3258	908
A2	3691	676
A3	3449	817
A4	3698	674
B1	3331	1062
B2	3750	806
B3	3516	923
B4	3757	792
C1	2911	1024
C2	3411	756
C3	3147	922
C4	3429	752
D1	2978	1211
D2	3441	909
D3	3155	1101
D4	3451	906
CURRENT DP	1783	1465

FIGURE 13

CURRENT DP VS SCHEMES A4,B4,C4,D4

10000 Drivers With Largest Estimated Accident Potential

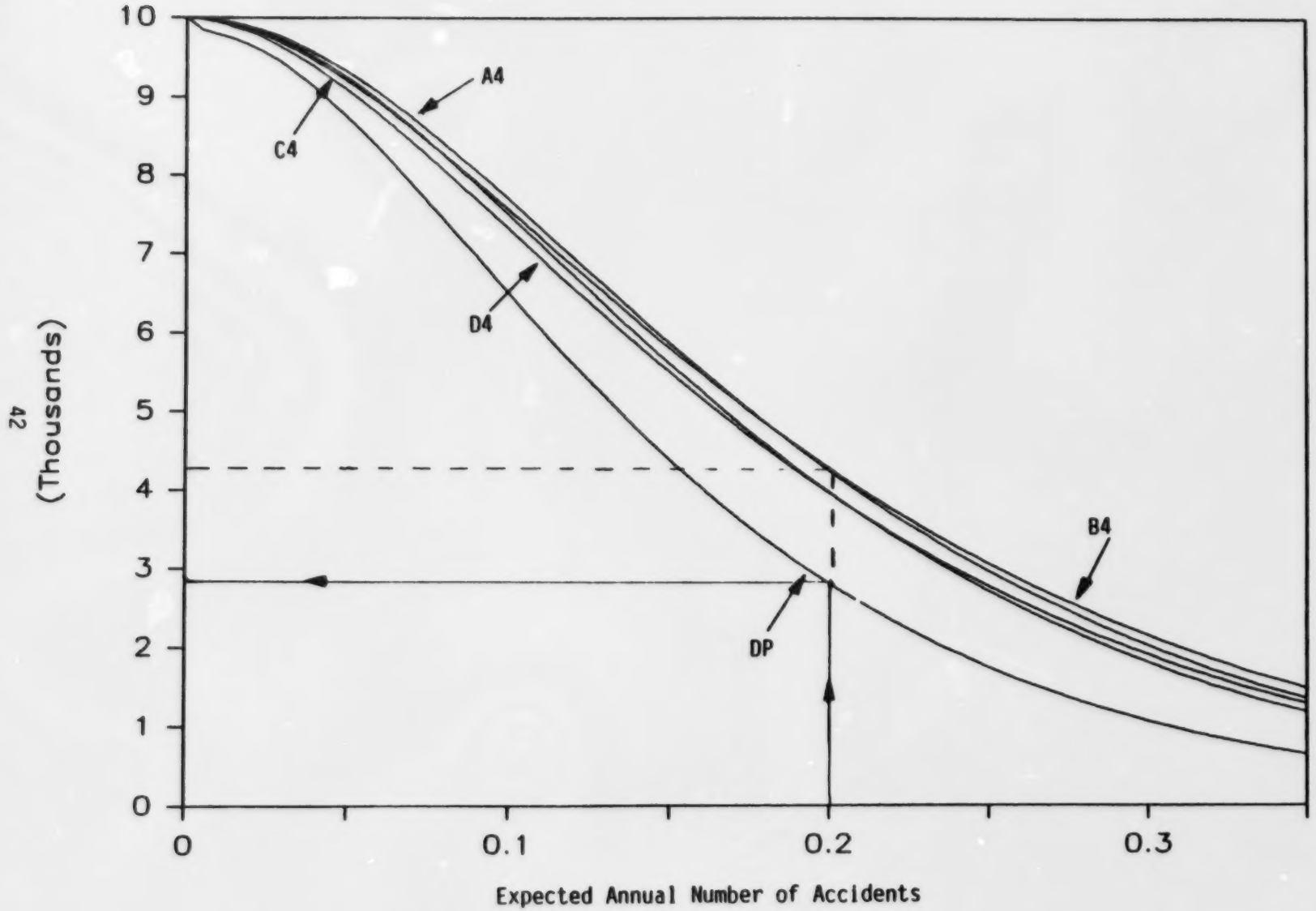
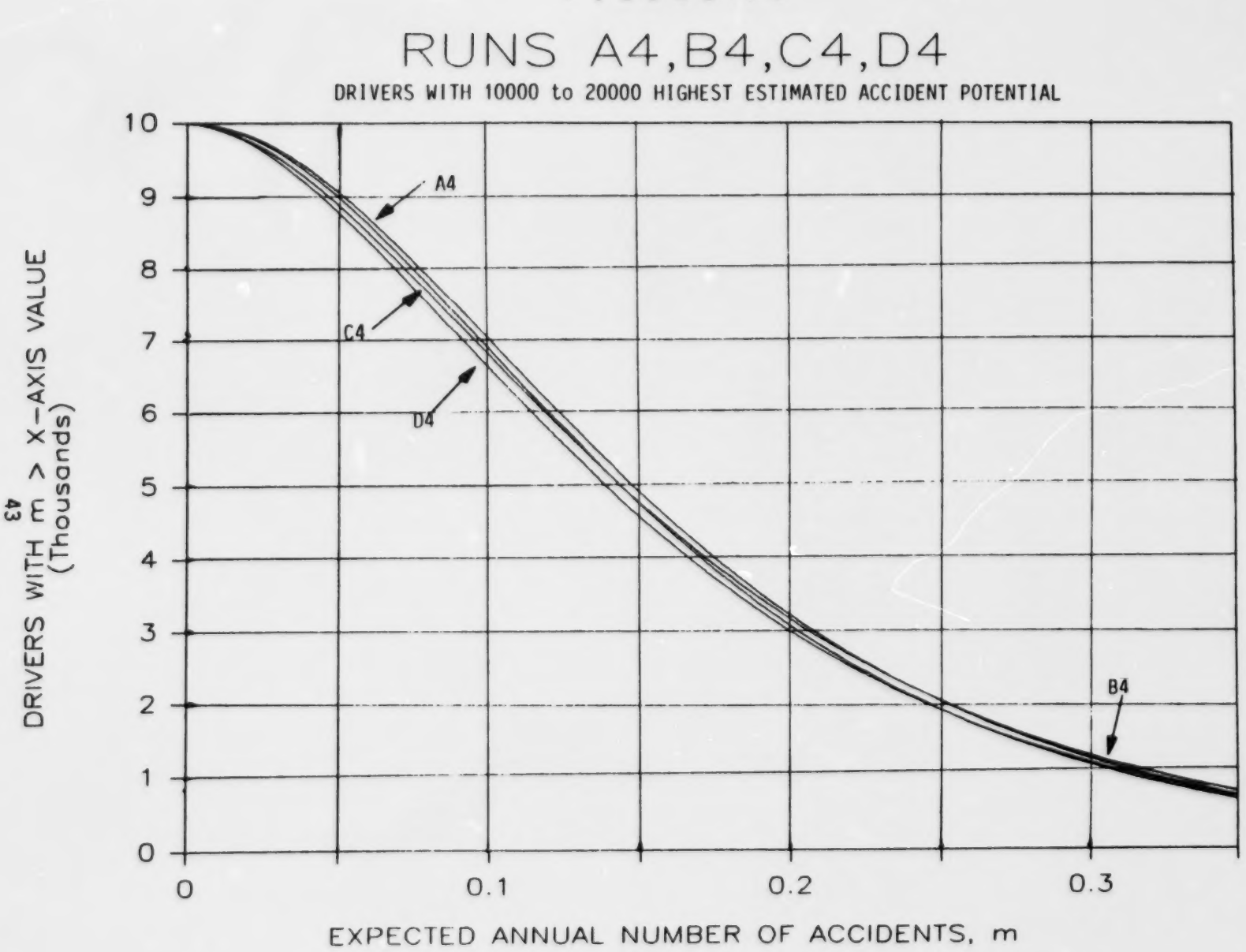


FIGURE 14


RUNS A4,B4,C4,D4
DRIVERS WITH 10000 to 20000 HIGHEST ESTIMATED ACCIDENT POTENTIAL
DRIVERS WITH m > X-AXIS VALUE
(Thousands)
43
10
9
8
7
6
5
4
3
2
1
0
A4
C4
D4
B4
0
0.1
0.2
0.3
EXPECTED ANNUAL NUMBER OF ACCIDENTS, m

5.0 SUMMARY AND CONCLUSIONS

The purpose of this work was to use a driver's record of convictions and accidents in order to predict, as well as possible, which drivers, based on their past record, are most likely to have an accident in the near future.

Driver Record Sample. A sample of 827,995 records of drivers licensed to drive in Ontario during 1981-1985 has been examined. Each driver record contained information about the driver's gender, age, and details of accidents, convictions, demerit points and suspensions.

Preparatory Analysis. In preparation for analysis, the many hundreds of offense types had to be grouped into a manageable number of categories. We first put in the same group offenses which were similar in nature and then consolidated those offenses which were associated with a similar average number of accidents. It turned out that, for example, drivers who in one year had a single speeding conviction had fewer accidents in the remaining three years than other drivers who had a single conviction in that year for a relatively minor offense such as a missing lamp. This finding may be initially puzzling but, on reflection, helps to interpret later results correctly. It arises partly because not all illegal behaviours lead to convictions at the same rate, partly because some offenses are specific to truck drivers who drive 10-20 times as much as car drivers and tend to have proportionately more accidents, and partly because behaviour which results in a fairly innocuous offense such as a noisy muffler may be of the type which also leads to accidents. Therefore neither the ratio of accidents to convictions nor the "weights" which are later attached to particular offenses are an indication of the gravity of that offense.

Drivers whose license is suspended will curtail their driving to some extent. This is why, during the period of suspension, one should expect some reduction in the number of accidents in which they are involved. However, the extent to which driving is curtailed is unknown. To assume that suspended drivers stop driving would introduce a bias into the analysis; assuming that they continue to drive would cause another bias. To protect the integrity of the results, drivers who were suspended had to be removed from the analysis. Therefore, our conclusions apply directly only to drivers who have not been suspended. The extension to drivers who were suspended under the present demerit point system is therefore an extrapolation.

Alternate Weighting Schemes. We have examined 16 models to estimate the expected number of future accidents for a driver based on age, gender, convictions and accidents. The models differ from each other in the information which is used. Some make use of age and gender, others do not; in some, each of 14 types of convictions is given a different weight, in others all convictions have the same weight; in some, at-fault accidents are counted separately from not-at-fault accidents, in some they are lumped together. All 16 models have a common structure, that of a "weighed sum" of convictions, and in some models, accidents. The basis is always a 17-20 year old male with no accidents or convictions in the first two years. Thus, in one of the models, these drivers are expected to have 0.176 accidents in the second two year period. For being female one subtracts 0.061, for being 24 years old one subtracts 0.039, for having two speeding convictions one adds 2 x 0.027 for having an accident one adds 0.058, and so on. The result is an estimate of the accident potential of a driver with specific age, gender and

record of convictions and accidents. We also estimate the distribution of accident potential in the population of drivers with given traits. We make use of this distribution to describe the performance of alternative models.

Measures of Performance. Two measures of performance were used to judge the quality of a model or weighting scheme. The first measure of performance is the *number of period two accidents* in a group of drivers of fixed size as identified by each model or weighting scheme on the basis of their first period record. The second measure of performance is the number of "hits" and the number of "false alarms" expected to occur in groups of fixed size as selected by the alternative models.

Comparison of the Current Demerit Point System and Alternative Schemes. By the first measure of performance we find that those 10,000 drivers who in the first two year period had most demerit points, recorded during the second two year period 1452 accidents per year. Those 10,000 drivers who in the first two year period had the most accidents, recorded in the subsequent two year period 1828 accidents per year. A third group of 10,000 drivers, those who by model A2 had the highest accident potential when calculated on the basis of first period data, recorded 2116 accidents per year in the second period. Thus, selection by model A2 gives a richer catch than either selection by previous accidents or by the current demerit points. On this score, model A2 seems to perform best.

Several conclusions emerge. First, one can do better than to use the current demerit point weights. Second, it is important to make use of the driver's record of accidents. Third, the more drivers that are identified, the lesser the "yield". Thus, the top 1000 drivers have an accident rate of ~0.3 accidents/year which is approximately 6 times the population average; for the first 10,000 drivers the average accident rate is ~ 0.2 and so on.

We now turn to the examination of the second measure of performance. Not all drivers have the same expected number of accidents per year. Some drive more, some less, many drivers are prudent, some take unwise risks. Based on the accident data we show how many drivers in Ontario have what expected number of accidents. Thus, for example, of 5 million drivers, some 90,000 have an expected number of accidents three standard deviations above the average for the population. It is these "high-accident-potential" drivers which a demerit point scheme aims to identify. A two year record of convictions and accidents is just too short to estimate a driver's expected number of accidents with accuracy. This is because a given two year period may not truly reflect a driver's typical driving behaviour. This is why some of those identified by the model as having the highest expected number of accidents, turn out in reality to be just average drivers. Conversely, this is why most high accident potential drivers may not have, in two years, a record which identifies them as such. We showed, for example, that of the 10,000 drivers who by the "richest" model (A4) were estimated to have the highest expected number of accidents, 3698 (Table 12) have an accident potential in excess of three standard deviations above the mean for the population. These are the hits. At the same time, 674 of those 10,000 were average drivers or better. Were we to identify a different group of 10,000 drivers, those who in a two year period had the most demerit points (using the current scheme), we would have found 1783 hits (Table 11), instead of 3698 by model A4, and 1465 false alarms, instead of 674 by model A4. Thus, the model A4 is considerably better than the current demerit point

instead of 3698 by model A4, and 1465 false alarms, instead of 674 by model A4. Thus, the model A4 is considerably better than the current demerit point scheme. It identifies more high-accident-potential drivers and fewer drivers who are better-than-average.

Since A4 uses age and gender data, it is not fair to compare it to the current demerit point scheme. A fair comparison is with model B1 (no age, gender or accident information used) or B2 (accident information used but no age or gender). The performance of these three models for the 10,000 drivers with highest estimated accident potential is shown below:

	Hits >0.22	False Alarms <0.055 acc./yr.
Current Demerit Points	1783	1465
B1	3331	1062
B2	3750	806

Comparison of Equal Point Schemes and Separate Weight Schemes. Finally it may be of interest to note that little is gained by giving different numbers of points to different offenses. Model D1 uses simply one-point-per-conviction and no accident data; model D2 uses one-point-per-conviction and 1.88 points per accident. Model D2 is only slightly worse than model B2 which assigns different numbers of points to each of fourteen offense classes.

	Hits >0.22	False Alarms <0.055 acc./yr.
Current Demerit Points	1783	1465
D1	2978	1211
D2	3441	909

In summary, if the purpose of a demerit point system is to identify drivers who are most likely to have an accident, the scheme used now is not as efficient as alternative schemes would be. Even by giving equal weights to all convictions and a weight of ~2 to an accident (D2), one can do much better. It is important to use data about accident involvement, but it does not pay to differentiate between at-fault and not-at-fault accidents.

For the top 5000 (or so) drivers, it makes little difference whether separate weights are assigned to different offenses. However, for the next 100,000 drivers, separate weights improve performance (in terms of predicting the number of future accidents) by a few percentage points. Separate weights help to increase the number of hits and to reduce the number of false alarms by some 10%.

Comparison of Schemes Using Age and Gender and Schemes That Do Not. For the top 5000 or so drivers, the inclusion of age and gender information appears to be unimportant. For the next 100,000 drivers consideration of age and gender improves performance (in terms of predicting future accidents).

Limitations of any Demerit Point Scheme. With all this, one has to keep in mind that if only a few drivers are identified (say about 10,000) 30-40% of those will be genuine high-accident-potential drivers and only 6-10% will be falsely identified better-than-average drivers. However, only 3% of all high-accident-potential drivers in the population will be in this group of 10,000 drivers. It does not help much to increase the size of the group because performance deteriorates with size. Thus, in a group of 120,000 drivers only ~19% are genuinely high-accident-potential whereas 13-16% are falsely identified. Even when as many as 120,000 drivers are identified by the richest model, only 22,363 of the 90,000 "high-accident-potential" drivers are caught in the net.

Further Improvements. The performance of models for the estimation of a person's accident potential can be further improved. We think that consideration should be given to a system which nearly continuously tracks a person's accident potential. If during a certain period of time (measured in weeks) the driver did not acquire a conviction and was not involved in an accident, his or her estimated accident potential would be revised downward. If during that period of time, convictions or accidents were recorded, the estimated accident potential would be correspondingly revised upward. A person's aging, the general accident trend and seasonal variation would also be reflected in these revisions. In this manner, a person's current estimated accident potential could be made a reflection of his or her entire driving history. In such a scheme, there is no need to specify an arbitrary period of time after which points are forgiven.

In the models developed so far, involvement in an accident adds a fixed amount to a driver's accident potential. Under the newly suggested scheme, an accident for a person with an already high accident potential would be weighted more heavily. In general a "revision" scheme of this nature relies on solid mathematical logic and is expected to perform better than other possible schemes.

APPENDIX 1

Recalculation of the Estimated m Associated With Each Offense Type According to a Standardized Age-Gender Distribution

Recalculation of the Estimated m Associated With Each Offense Type According to a Standardized Age-Gender Distribution

The clustering of all possible offense types into a manageable number of groups was accomplished in several steps. The first step was to combine the offense types which are quite similar in nature into a smaller number of categories. A total of 45 such categories were established in our discussions with MTO in April 1987. These consisted of 38 categories of moving violations, 3 for vehicle related offenses, and 4 for non-moving, administrative violations.

This number of categories was still considered too large for practical purposes. In addition it was expected that some categories would still contain too few convictions. Thus an attempt was made to further group similar categories. This time, the similarity or difference of convictions was indicated by the contribution of each conviction in explaining a driver's accident record. Some re-shuffling of convictions within the original categories was also done on this basis.

The object of the exercise was to identify drivers with one and only one conviction of a given type (or category) in one calendar year, and for each conviction type, obtain the mean (m) and the [var(m)] of accidents in the other 3 years. These estimates were adjusted so that the age-sex distribution of drivers in each conviction class was the same as that for all drivers who had one conviction (of any type) in any of the 4 years 1981-1984.

Step 1: Drivers with one and only one conviction (*of any type*) in any year were identified and grouped according to age and sex. This was designated the base population. The proportion of drivers in each age-sex group was then calculated. Table A1.1 shows, for example, that 8154 such drivers were males aged 30 at the time of the conviction, accounting for 0.0323 of the base population. The number 0.0323 is the weight that was applied in step 2.

Step 2: Drivers with one and only one conviction (*of a given type*) in any year were identified and grouped according to age and sex. For each group, the proportion and number of drivers, the number and sum of squares of accidents in the other 3 years was calculated. For example, Table A1.2 shows that of the 8154 30-year old male drivers (Table A1.1) having only one conviction in any year, 332 had running red light convictions. These 332 drivers represent 0.02595 of the sub-population of drivers whose one conviction in a year was Type m13, and had 99 accidents (Sum of squares = 149) in the other 3 years.

Step 3: For a given conviction type the number of accidents and sum of squares for each age-sex group were weighted and aggregated as follows. For an age-sex group, the ratio of the population weight to the sub-population weight was applied to number of accidents and sum of squares for the group. Thus, for 30 year old males, the ratio is 0.323/0.2595 = 1.244 for "run red light" convictions, shown in column 4 of Table A1.2; the number of accidents (99) and the sum of squares (149) are "factored" by 1.244. The factored numbers for each age group are summed and the sums are divided by sub-population size to get the (weighted) mean values for the sub-population. Table A1.3 shows that of all drivers in the "base" population, 13731 had "run red light" convictions and had 4270 accidents in 3 years. The method

described above gives a weighted mean of 0.313 accidents per driver [E(m)] in this class.

Step 4: 95% confidence limits for the estimates of E(m) were calculated. These are also shown in Table A1.3.

Step 5: Steps 2 to 4 were repeated for speeding convictions placed into categories according to the following guidelines:

(a) Reflecting the current demerit system, drivers were grouped regardless of speed limit according to whether they had exceeded the speed limit by less than 16 km/h, 16-29 km/h, 30-49 km/h or more than 49 km/h.

(b) For drivers grouped according to (a), sub-groups were identified for speed-limit groups of 20-40 km/h, 50-60 km/h and 80-100 km/h.

Table A1.4 provides similar information to Table A1.3 but for the various types of speeding convictions.

TABLE A1.1: Distribution of drivers with one conviction in any year

	MALES				FEMALES			
Age	Drivers	Proportion	Accs.	Sum of Squares	Drivers	Proportion	Accs.	Sum of Squares
17	1594	0.006309	534	961	260	0.001029	56	73
18	4534	0.017946	1810	3990	961	0.003804	186	306
19	7712	0.030525	3075	8182	1691	0.006693	395	699
20	10034	0.039716	4247	12502	2353	0.009313	527	896
21	11253	0.044541	4572	14375	2766	0.010948	614	1505
22	11558	0.045748	4734	13799	2879	0.011395	628	1144
23	11253	0.044541	4238	13090	3034	0.012009	615	1422
24	10740	0.042510	4019	12121	2899	0.011475	572	952
25	10254	0.040586	3474	10120	2873	0.011372	560	1050
26	9756	0.038615	3402	9553	2743	0.010857	503	971
27	9284	0.036747	3065	8657	2670	0.010568	449	849
28	8846	0.035013	2775	8016	2583	0.010224	432	796
29	8361	0.033094	2534	7125	2669	0.010564	520	981
30	8154	0.032274	2479	7070	2570	0.010172	461	827
31	7895	0.031249	2371	6021	2537	0.010042	426	674
32	7013	0.027758	2079	5416	2463	0.009749	437	846
33	6126	0.024247	1849	4759	2240	0.008866	430	797
34	5143	0.020357	1519	4110	2185	0.008648	351	678
35	4583	0.018140	1352	3397	1969	0.007794	374	747
36	3762	0.014890	1175	3013	1663	0.006582	309	551
37	3376	0.013363	976	2702	1589	0.006289	288	609
38	2986	0.011819	912	2716	1313	0.005197	260	537
39	2679	0.010604	858	2500	1172	0.004639	228	407
40	2366	0.009365	687	1762	1017	0.004025	187	338
41	2252	0.008914	647	1574	935	0.003701	155	320
42	2024	0.008011	610	1596	850	0.003364	147	235
43	1814	0.007180	479	1150	792	0.003135	131	189
44	1709	0.006764	516	1260	597	0.002363	100	166
45	1530	0.006056	410	875	622	0.002462	122	185
46	1380	0.005462	406	888	537	0.002126	81	129
47	1312	0.005193	384	833	465	0.001841	80	148
48	1134	0.004488	337	875	454	0.001797	83	118
49	1086	0.004299	333	952	404	0.001599	75	130
50	951	0.003764	283	905	320	0.001267	50	85
51	882	0.003491	244	640	318	0.001259	44	61
52	846	0.003349	200	499	295	0.001168	46	46
53	762	0.003016	209	428	252	0.000997	42	65
54	663	0.002624	233	767	297	0.001176	47	76
55	617	0.002442	164	365	233	0.000922	59	98
56	565	0.002236	174	356	207	0.000819	31	62
57	492	0.001947	133	269	188	0.000744	31	44
58	460	0.001821	98	185	173	0.000685	43	79
59	376	0.001488	98	234	138	0.000546	29	58
60	372	0.001472	106	221	145	0.000574	32	52

61	297	0.001176	78	168	114	0.000451	16	33
62	295	0.001168	91	226	94	0.000372	12	19
63	221	0.000875	68	159	82	0.000325	9	11
64	177	0.000701	64	113	88	0.000348	15	27
65	159	0.000629	40	88	61	0.000241	5	8
66	140	0.000554	31	57	55	0.000218	10	16
67	113	0.000447	37	154	52	0.000206	13	22
68	104	0.000412	20	44	43	0.000170	7	13
69	69	0.000273	10	16	40	0.000158	9	22
70	73	0.000289	27	57	25	0.000099	10	14
71	61	0.000241	12	15	18	0.000071	1	1
72	56	0.000222	20	29	19	0.000075	2	2
73	26	0.000103	9	15	10	0.000040	6	10
74	40	0.000158	21	63	18	0.000071	0	0
75	19	0.000075	2	2	18	0.000071	9	13
76	38	0.000150	12	33	14	0.000055	2	2
77	36	0.000142	14	167	7	0.000028	1	1
78	29	0.000115	11	75	10	0.000040	3	5
79	20	0.000079	4	9	11	0.000044	3	3
80	30	0.000119	6	14	4	0.000016	2	5
81	21	0.000083	8	12	1	0.000004	1	1
82	14	0.000055	5	7	3	0.000012	0	0
82+	9	0.000036	1	1	2	0.000008	0	0

TABLE A1.2: Drivers with 1 "run red light" conviction in any year

	MALES					FEMALES				
		Proportion					Proportion			
Age	No.	For class	Relative to all	Accs.	Sum Sq.	No.	For class	Relative to all	Accs.	Sum Sq.
17	84	0.00657	0.961	21	21	15	0.00117	0.878	4	6
18	247	0.01931	0.929	91	121	66	0.00516	0.737	13	13
19	422	0.03299	0.925	159	244	121	0.00946	0.708	24	32
20	497	0.03885	1.022	183	293	130	0.01016	0.916	28	34
21	571	0.04463	0.998	234	410	167	0.01305	0.839	27	33
22	546	0.04268	1.072	195	295	202	0.01579	0.722	31	39
23	501	0.03916	1.137	150	211	149	0.01165	1.031	15	17
24	492	0.03846	1.105	163	329	161	0.01259	0.912	18	20
25	436	0.03408	1.191	120	163	159	0.01243	0.915	30	36
26	428	0.03346	1.154	143	244	142	0.01110	0.978	16	18
27	374	0.02923	1.257	105	196	138	0.01079	0.980	25	25
28	375	0.02931	1.194	80	178	124	0.00969	1.055	22	26
29	328	0.02564	1.291	102	162	129	0.01008	1.048	24	34
30	332	0.02595	1.244	99	149	143	0.01118	0.910	25	31
31	313	0.02447	1.277	57	75	112	0.00875	1.147	16	16
32	298	0.02329	1.192	83	115	143	0.01118	0.872	21	25
33	261	0.02040	1.188	66	98	131	0.01024	0.866	29	43
34	259	0.02025	1.006	67	90	110	0.00860	1.006	10	10
35	220	0.01720	1.055	59	79	108	0.00844	0.923	15	15
36	188	0.01470	1.013	61	111	103	0.00805	0.818	14	14
37	185	0.01446	0.924	47	63	105	0.00821	0.766	11	11
38	152	0.01188	0.995	44	76	83	0.00649	0.801	19	21
39	147	0.01149	0.923	41	55	59	0.00461	1.006	4	4
40	139	0.01087	0.862	44	72	82	0.00641	0.628	12	12
41	117	0.00915	0.975	27	35	58	0.00453	0.816	8	12
42	105	0.00821	0.976	18	22	66	0.00516	0.652	5	5
43	101	0.00789	0.909	27	45	61	0.00477	0.657	12	14
44	81	0.00633	1.068	18	22	39	0.00305	0.775	8	10
45	102	0.00797	0.760	21	25	56	0.00438	0.562	13	21
46	89	0.00696	0.785	18	20	36	0.00281	0.755	1	1
47	79	0.00618	0.841	24	34	38	0.00297	0.620	4	6
48	67	0.00524	0.857	17	33	41	0.00320	0.561	7	7
49	52	0.00406	1.058	9	11	34	0.00266	0.602	5	5
50	54	0.00422	0.892	17	25	28	0.00219	0.579	6	10
51	56	0.00438	0.798	13	17	28	0.00219	0.575	3	3
52	63	0.00492	0.680	11	17	22	0.00172	0.679	5	5
53	42	0.00328	0.919	7	7	17	0.00133	0.750	2	2
54	42	0.00328	0.799	17	31	33	0.00258	0.456	2	2
55	44	0.00344	0.710	10	16	20	0.00156	0.590	4	6
56	30	0.00235	0.954	7	7	15	0.00117	0.698	2	2
57	36	0.00281	0.692	3	3	18	0.00141	0.529	5	5
58	30	0.00235	0.777	6	6	10	0.00078	0.876	4	8
59	30	0.00235	0.635	8	8	9	0.00070	0.776	0	0
60	31	0.00242	0.607	8	20	9	0.00070	0.816	1	1

61	16	0.00125	0.940	5	7	12	0.00094	0.481	3	5
62	28	0.00219	0.534	3	3	7	0.00055	0.680	1	1
63	20	0.00156	0.560	4	6	6	0.00047	0.693	0	0
64	12	0.00094	0.747	3	3	4	0.00031	1.113	0	0
65	13	0.00102	0.619	2	2	6	0.00047	0.514	0	0
66	12	0.00094	0.591	4	6	2	0.00016	1.394	1	1
67	8	0.00063	0.715	0	0	3	0.00023	0.878	1	1
68	11	0.00086	0.479	1	1	2	0.00016	1.087	0	0
69	2	0.00016	1.746	0	0	3	0.00023	0.674	0	0
70	5	0.00039	0.739	0	0	2	0.00016	0.633	1	1
71	7	0.00055	0.440	2	2	2	0.00016	0.454	0	0
72	5	0.00039	0.568	0	0	2	0.00016	0.480	1	1
73	2	0.00016	0.659	0	0	0	0.	0.480	0	0
74	1	0.00008	2.021	1	1	1	0.00008	0.908	0	0
75	1	0.00008	0.959	0	0	0	0.	0.908	0	0
76	5	0.00039	0.384	1	1	4	0.00031	0.176	1	1
77	4	0.00031	0.454	0	0	3	0.00023	0.119	0	0
78	7	0.00055	0.210	3	5	0	0.	0.119	0	0
79	1	0.00008	1.011	2	4	2	0.00016	0.281	0	0
80	3	0.00023	0.507	0	0	0	0.	0.281	0	0
81	1	0.00008	1.062	0	0	0	0.	0.281	0	0
82	1	0.00008	0.704	1	1	1	0.00008	0.154	0	0

TABLE A1.3: Accidents for Drivers with 1 Conviction in 1 Year

Category	Brief Description	No. of Drivers	3 yr. Accs.	Weighted Mean	95% Limits Upper	Lower
n1	Minor neglect, licenses, permits	6495	2918	0.434	0.445	0.422
n2	Neglect, insurance, permits, etc.	1589	719	0.414	0.438	0.392
n3	License suspended, HTA	874	454	0.424	0.456	0.394
n4	Learners	34	18	0.343	0.502	0.212
v1	Minor veh.; lamps, noise	2954	1498	0.468	0.485	0.451
v2	Brakes, tires, unsafe vehicle	946	451	0.400	0.430	0.371
v3	Comm. veh; size & weights	503	369	0.542	0.583	0.500
m1	Seat belt	12337	4858	0.376	0.384	0.368
m2	Speeding	173592	55211	0.319	0.321	0.317
m3	Careless driving	902	342	0.327	0.357	0.299
m4	Slow driving	45	11	0.119	0.237	0.055
m8	STOP sign, ROW violations	14024	3935	0.288	0.295	0.281
m9	PXO violations	1237	355	0.296	0.320	0.272
m10	Turns violations; right, left, U	18231	4942	0.283	0.289	0.277
m11	Unsafe move; open door	1649	542	0.334	0.355	0.312
m13	Disobey red light	13731	4270	0.313	0.321	0.306
m14	Amber light	3453	982	0.285	0.299	0.271
m15	Advance green	274	73	0.265	0.317	0.218
m16	Fail to share road	170	62	0.303	0.372	0.242
m17	Passing violations	1305	459	0.327	0.351	0.303
m18	Wrong-way one way street	1582	458	0.284	0.306	0.264
m19	Improper driving divided h'way	2599	900	0.361	0.379	0.344
m20	F.T.C.	934	337	0.344	0.374	0.316
m21	Emerg. veh., school x'ing	48	15	0.159	0.280	0.084
m22	R/R crossing violations	95	35	0.314	0.408	0.233
m24	Headlight beam not lowered	225	71	0.260	0.318	0.209
m25	Improper parking	145	77	0.407	0.484	0.334
m26	Fail stop for school bus	604	133	0.281	0.317	0.249
m28	Disobey traffic signs	1650	529	0.322	0.344	0.301
m29	Fail report accident	224	73	0.266	0.324	0.215
m30	Fail remain at scene	236	75	0.315	0.373	0.261
m32	Dangerous driving C.C.C.	5	2	0.010	0.421	0.000
m33	Fail remain at accident C.C.C.	66	39	0.377	0.491	0.274
m34	Dangerous driving C.C.C.	89	46	0.281	0.376	0.202
m35	Impaired driving C.C.C.	2381	1040	0.443	0.462	0.424
m36	Fail/refuse breath test C.C.C.	94	37	0.218	0.306	0.149
m37	Fail or ref. breath test C.C.C.	120	50	0.314	0.397	0.241
m38	Driving with >80 mgs. alcohol	3676	1502	0.386	0.401	0.371
m41	Crowding driver seat	128	36	0.280	0.362	0.211
m44	Radar device in vehicle	68	27	0.251	0.359	0.166
m45	No safe helmet, motorcycle	96	47	0.266	0.357	0.191
m46	Fail to signal to stop	18	6	0.080	0.277	0.019
m47	FTC, commercial vehicle	67	35	0.362	0.475	0.262
m48	Fail to stop for police officer	12	6	0.086	0.340	0.016

TABLE A1.4: Accidents for drivers with 1 speeding conviction in 1 year

Speed Limit	Above Limit	No. of Drivers	3 yr. Accs.	Weighted Mean	95% Limits Upper	Lower
All	Ø-15	123224	33755	Ø.293	Ø.296	Ø.291
All	16-29	79589	22479	Ø.286	Ø.289	Ø.283
All	3Ø-49	23891	7331	Ø.297	Ø.3Ø2	Ø.291
All	>49	894	371	Ø.315	Ø.344	Ø.286
2Ø-4Ø	Ø-15	24264	63Ø7	Ø.288	Ø.293	Ø.283
5Ø-7Ø	Ø-15	8Ø248	22Ø37	Ø.294	Ø.297	Ø.291
8Ø-1ØØ	Ø-15	18712	5411	Ø.292	Ø.298	Ø.286
2Ø-4Ø	16-29	5276	1413	Ø.281	Ø.293	Ø.27Ø
5Ø-7Ø	16-29	39995	11427	Ø.293	Ø.297	Ø.288
8Ø-1ØØ	16-29	34318	9639	Ø.281	Ø.285	Ø.276
2Ø-4Ø	3Ø-49	683	185	Ø.269	Ø.3Ø2	Ø.239
5Ø-7Ø	3Ø-49	93Ø4	2914	Ø.3Ø5	Ø.314	Ø.296
8Ø-1ØØ	3Ø-49	13Ø4	4232	Ø.292	Ø.3ØØ	Ø.285
2Ø-4Ø	>49	14	8	Ø.186	Ø.432	Ø.Ø64
5Ø-7Ø	>49	4Ø5	183	Ø.341	Ø.386	Ø.299
8Ø-1ØØ	>49	475	18Ø	Ø.272	Ø.312	Ø.236

APPENDIX 2

Number of Convictions in Each of the Final 14 Conviction Categories

Appendix 2

1981-1984 CONVICTIONS BY PROPOSED CONVICTION CODE CATEGORIES
(Based on sample of 8000 drivers with record)

Offence	Code	TOTAL CONVICTIONS IN 1981	1982	1983	1984	OLD ID	Tot.	Acc Pts.
Total for all convictions		**3588**	**3842**	**3617**	**3270**	**14209**	**976**	
REVISED CATEGORY nn1								
Dr. m/v,no valid permit	10033	0	13	46	68	n1	127	2
fail not. ch. address	10050	0	9	5	5	n1	19	3
Fail rep.dam to h'wy prop.	12760	1	1	1	1	n1	4	3
" " " " " "	12770	1	0	0	1	n1	2	2
Fail to have insurance card	20050	0	0	23	34	n1	57	3
" " carry " "	20060	0	0	3	11	n1	14	
" " " evid. of ins.	20070	0	0	2	0	n1	2	
" " fail not. addr. ch.	90810	0	7	6	3	n1	16	3
No driver's license	10190	26	54	26	22	n2	128	10
Driv., cond. proh.	10193	0	5	4	5	n2	14	2
Learner unaccompanied	90700	1	0	0	0	n4	1	
Class R Drive m/c at night	90710	0	1	1	0	n4	2	
" " Cont. to lic.cond.	90790	1	0	0	0	n4	1	
REVISED CATEGORY nn2								
Permit unlic. driv.	10195	5	1	1	0	n2	7	
No lic., fail to ID self	10214	0	1	1	0	n2	2	
More than 1 license	13270	0	0	1	1	n2	2	
No insurance	20010	0	0	2	1	n2	3	
Op. m/v, no insurance	20011	0	0	3	5	n2	8	3
False statem. re. insurance	20030	0	0	2	0	n2	2	
Alter/deface no. plate	10100	0	0	1	0	n2	1	
Allow/use altered plates	10110	0	2	0	1	n2	3	
Plates unlawf. removed	10120	0	0	1	0	n2	1	
Pos. canc/susp. license	10222	0	0	1	0	n2	1	
Lend driver license	10224	0	0	1	0	n2	1	
Fail surr. dr. license	10228	0	1	0	0	n2	1	
Unlaw. pos. license	10320	1	1	0	0	n2	2	
Dr. m/v, no valid plates	10034	0	0	2	7	n1	9	
No valid. tag on plates	10035	0	0	5	34	n1	39	
Draw trailer, no permit	10037	0	0	12	22	n1	35	
fail surr. m/v permit	10038	0	0	3	15	n1	18	
" " trl. "	10039	0	0	0	2	n1	2	
No plate	10060	0	1	0	0	n1	1	
No plate, operate m/v	10063	0	29	5	1	n1	35	
No val. tag on m/v	10067	0	13	2	0	n1	15	

Appendix 2

Imp. plate pos. m/v	10070	0	1	0	1	n1	2	
Fail apply for permit	10071	0	0	2	1	n1	3	
Draw trl., no plate	10073	0	1	0	0	n1	1	
Imp. plates on veh.	10130	0	0	2	11	n1	13	
Use unauthor. plate val.	10136	0	0	0	1	n1	1	
Expose oth. no. near plate	10160	0	1	0	0	n1	1	
No. on plate not visible	10170	0	2	0	0	n1	2	
Fail to prod. dr. license	10200	0	0	0	1	n1	1	
" " " " "	10210	0	89	38	20	n1	148	9
No name on comm. veh.	11060	0	0	1	0	n1	1	
Draw trailer, no plate	12791	0	0	0	2	n1	2	
Fail to carry/prod.permit	13210	0	4	2	0	n1	6	1
Fail ret.pl. conv. veh.	90234	0	4	0	1	n1	5	
Dealer plate infraction	90250	0	1	0	0	n1	1	
Fail sign license	90820	0	3	0	2	n1	5	
Dr. lic. susp. H.T.A.	10332	39	22	25	10	n3	96	5
" " " "	10335	0	0	1	0	n3	1	

REVISED CATEGORY vv1

No sl-mov-veh sign	10910	0	0	0	1	v1	1	
Improper trailer attachment	10960	3	3	3	1	v1	10	
Imp./insuff./ drvg. lamps	10420	0	21	24	20	v1	65	2
Imp. lights, motor cycle	10423	0	1	2	0	v1	3	
Imp. head/drvg. lamps	10440	0	1	0	0	v1	1	1
Cov. or coat lamps	10445	0	1	1	1	v1	3	
Imp. ident. lamps	10480	0	0	0	1	v1	1	
Imp. red front lamp	10500	0	1	1	0	v1	2	
Imp. no. plate lamp	10530	0	5	5	6	v1	16	
Imp. park lights	10540	0	0	1	1	v1	2	
Imp. rear lamp on trl.	10570	0	4	3	2	v1	9	
Imp. rr. lamp, obj.>leg. w.	10580	0	0	1	0	v1	1	
Imp. signal devices	10610	0	5	1	0	v1	6	
No mudguards	10730	0	1	1	0	v1	2	
View obstr., sign/poster	10800	0	1	0	0	v1	1	1
" " , articles hung	10810	1	0	0	0	v1	1	
Obscure color window	10815	0	4	2	2	v1	8	
Windows obstructed	10830	2	0	0	0	v1	2	
" "	10840	3	4	1	0	v1	8	1
Rear window obstructed	10850	1	2	0	1	v1	4	
Def/imp/no muffler	10860	0	2	6	5	v1	13	
Proh. use studded tires	90520	0	0	0	1	v1	1	

REVISED CATEGORY vv2

Def./no brakes, m/v	10650	1	0	1	0	v2	2	
" " " , trailer	10670	0	0	0	1	v2	1	
Def. brakes, contr. to reg.	10680	0	2	2	2	v2	6	2
Def./imp tires	10750	0	1	4	3	v2	8	

Appendix 2

Use proh./unsafe veh	10970	0	1	0	0	v2	1	
Refuse exam. unsafe veh.	10980	4	3	3	4	v2	14	
Drive unsafe veh.	10990	10	6	3	2	v2	21	3
Rm./mod./inop. s. belt ass.	11095	0	5	3	3	v2	11	1
REVISED CATEGORY vv3								
Overweight dual-axle	13060	0	2	0	2	v3	4	
Gross veh overweight	13150	0	1	0	1	v3	2	
Overweight during freezeup	13180	0	0	1	0	v3	1	
Over weight excess of perm	13200	0	1	2	3	v3	6	
Insecure load	11260	0	5	5	3	v3	13	
REVISED CATEGORY vv4								
Op. veh., fl. displ. dev.	10993	0	1	0	0	v3	1	
Imp. war. eq., dis. comm./v	12420	0	1	0	0	v3	1	
Not mark overhang load	11250	0	0	1	0	v3	1	
Excess veh. width	11310	0	0	1	0	v3	1	
Excess len. veh-comb.	11320	0	1	1	1	v3	3	
Excess veh. height	11350	0	1	2	0	v3	3	
REVISED CATEGORY ee1								
Excess fumes/smoke	10870	0	2	2	2	v1	6	
Unnecess. noise	10880	0	39	19	16	v1	74	
Horn/bell violation	10890	0	2	1	2	v1	5	
By-law unnec. noise	80030	0	0	0	1	v1	1	
REVISED CATEGORY mm1								
Fail yield ROW	11560	3	4	0	0	0	m8	4	4
Disobey police signal	11530	3	0	2	2	1	m8	5	0
Disobey stop sign	11570	3	0	1	1	1	m8	3	
Fail to stop, intersec.	11580	3	146	138	127	78	m8	489	20
Fail to yield ROW at stop	11590	3	26	26	18	15	m8	86	67
Fail stop at intersec.	11592	3	0	0	0	12	m8	12	2
Fail yield ROW at stop	11593	3	0	0	0	4	m8	4	3
Fail yield ROW at yield	11610	3	4	0	4	2	m8	10	7
Fail to yield ROW, Pr. rd.	11620	3	18	23	17	10	m8	68	54
Fail yld. ROW at inters.	11875	3	0	0	0	1	m8	1	1
Fail yield ROW	11910	3	2	4	0	2	m8	8	
Imp. RT at intersection	11693	2	4	7	6	10	m10	27	4
Imp. RT multi-ln. h'way	11703	2	3	4	1	1	m10	9	1
Improper left turn	11713	3	19	12	22	19	m10	72	68
Imp. LT at intersection	11723	2	10	14	15	16	m10	55	7
Improper left turn	11733	2	2	2	4	0	m10	8	1
Proh. U-turn	11810	2	0	0	0	1	m10	1	1
Proh. U-turn, hill crest	11840	2	1	0	0	0	m10	1	

Appendix 2

Prohibited turn	11920	2	85	91	64	40	m10	280	
By-law, proh. U-turn	80080	2	1	3	1	1	m10	6	
U-turn unsafe	80015		0	0	1	0	m10	1	
Prohibited turn	80020	2	0	0	34	61	m10	95	
By-law, proh. Turn	80110	2	57	66	47	12	m10	182	
" amber "	11870	3	52	37	31	18	m14	138	3
Disobey amber light	11921	3	0	0	0	3	m14	3	1
Disob. green arrow	11900	3	1	1	2	0	m15	4	
Fail proc. at flash. green	11905	3	2	1	1	1	m15	5	
Wrong-way,one-way street	12110	3	17	17	10	14	m18	58	2
PXO viol. same side of rd.	11640	2	2	3	1	1	m9	7	1
" " oth. " " "	11650	2	1	2	2	1	m9	6	
" "	11660	2	6	4	0	2	m9	12	
Pass at Ped. crossing	11665	2	1	5	0	6	m9	12	
Fail stop for sch. bus. o/t	12500	4	2	1	0	0	m26	3	
" " " " " meet	12510	4	3	3	3	2	m26	11	
" " " " " o/t	12511	6	0	0	1	0	m26	1	

REVISED CATEGORY mm2

Disobey red light	11860	3	160	150	122	79	m13	511	46
Disob. flashing red	11880	3	5	1	2	6	m13	14	4
Disobey red light	11925	3	0	0	0	19	m13	19	2
Dis. portable red light	11926	3	0	0	1	0	m13	1	
Fail stop at r/r cross.	12250	5	0	0	1	0	m22	1	
Crossing r/way barrier	12260	3	1	1	1	0	m22	3	

REVISED CATEGORY mm3

Unnec. slow driving	11520	2	2	2	0	0	m4	4	1
Fail to stop for emer/veh.	12200		3	1	0	0	m21	4	
" obey sch. cr. stop sign	12535		0	1	0	0	m21	1	
Fail lower h'light beam app.	12325	2	1	1	2	0	m24	4	0
Fail to rep. accident	12710	3	11	13	13	9	m29	46	36
Crowding driver seat	12240	3	1	2	3	3	m41	9	1
Radar dev in m/v	10925		0	1	0	2	m44	3	
No safe helmet, motor cy.	11090		0	5	3	0	m45	8	
Fail to signal to stop	11780	2	0	0	1	0	m46	1	
Fail stop for pol. officer	12785		0	1	2	1	m48	4	
Crim neg. caus. death CCC	70030		0	0	1	0	m31	1	1
Dang. driving CCC	70060		0	0	0	1	m32	1	0
Danger. driving CCC	70120		2	4	1	3	m34	10	6
Careless driving	11490	6	55	47	45	46	m3	193	142

REVISED CATEGORY mm4

Unsafe move	11750	2	33	33	35	25	m11	126	75
" " from parked pos.	11760	2	6	11	11	9	m11	37	26
Imp. open., veh door	12270	2	0	1	0	1	m11	2	

Appendix 2

Fail to share road	11940	2	10	5	7	6	m16	28	16
Fail share when overtaken	11960	2	1	0	0	0	m16	1	
" " " overtaking	11970	2	2	1	1	0	m16	4	3
Fail to turn out to rt/lt	11980	2	0	1	0	0	m16	1	
Improper passing	12000	3	0	1	0	0	m17	1	
Imp. pass.- appr. traffic	12010	3	3	1	2	0	m17	6	1
Imp. pass. - overt.traffic	12020	3	0	0	1	1	m17	2	1
Driving left of center	12040	3	4	7	2	2	m17	15	1
Dr. l. of cen. lev. cross.	12045	3	0	0	1	0	m17	1	
Pass on right/off r'way	12100	3	8	9	8	4	m17	29	3
Imp. pass. stpd. str.car	12290	3	1	2	1	0	m17	4	
Disob. traffic sign	11885	2	0	0	0	4	m28	4	
Disobey legal sign	12610	2	10	17	10	11	m28	48	1
Fail st. sign or mk. int.	11851		0	0	0	1	m28	1	1
Disobey sign re- tunnels	12620		0	0	0	1	m28	1	
By-law, disob. sign	80160		0	1	4	6	m28	11	

REVISED CATEGORY mm5

Fail remain at scene accid.	12730	7	7	7	4	9	m30	27	17
Fail/ref. breath test CCC	70185		1	2	3	2	m36	8	
Fail or ref. breath sam.CCC	70212		3	8	6	6	m37	23	5
Fail rem at accid. CCC	70090		3	3	2	10	m33	18	11
Driving with >80 mgs.alcoh.	70214		63	62	46	66	m38	237	37
Impaired driving C.C.C.	70174		40	56	46	42	m35	184	41

REVISED CATEGORY mm6

Drive on closed h'way	11556	3	1	1	0	2	m19	4	
Imp. dr. multi-ln. h'way	11930		1	1	1	0	m19	3	
Fail obey paved shld. signs	12105		0	4	1	3	m19	8	
Imp. driving, div. h'way	12120	3	18	13	21	11	m19	63	31
Imp. use cen. ln.,3-lane r.	12130	3	0	0	0	1	m19	1	
Disobey off.signs,div.h'way	12140	3	14	18	12	7	m19	51	
Imp. driving,div.h'way	12150	3	1	2	4	2	m19	9	1
Dr. m/v on sidewalk	80016		0	0	0	2	m19	2	
Dr. in des. bus lane	80012		0	0	2	3	m19	5	
Enter proh. h'way/r'dway	80022		0	0	2	3	m19	5	
F.T.C.	12170	4	1	1	1	0	m20	3	2
F.T.C., motor vehicle	12180	4	41	35	44	54	m20	174	133
Imp. parking	12360		0	0	1	1	m25	2	
Imp. parking	12380		0	1	0	0	m25	1	
Imp. park., Interf. w.traf	12430		0	3	1	1	m25	5	
Improper parking	90510		0	0	0	2	m25	2	
F.T.C., commer. m/v	12190	4	1	1	1	1	m47	4	1
Dr. comm. veh. lt. lane	80010		0	0	1	0	m47	1	
Fail/imp. use of Seat Belt	11097		228	227	250	126	m1	831	22
Fail ens. passen. s. belt	11099		0	1	10	13	m1	24	2

Appendix 2

REVISED CATEGORY mm7

Speeding -- km in --km zn.	11355	6	2269	2212	2127	1949 m2	8557 4
Speeding	11485		Ø	Ø	1	1 m2	2

APPENDIX 3

Model Weights for Alternate Demerit Point Schemes

Appendix 3

Regression Run A1: Age, sex as dummies, no accidents

Description of variable	Estimated Coefficient	Standard Error
Intercept (Male, Age < 21)	0.1839	0.00284
Dummy variable for age 21-25	-0.03752	0.002832
Dummy variable for age 26-30	-0.0564	0.002894
Dummy variable for age 31-35	-0.0609	0.003025
Dummy variable for age 36-40	-0.05246	0.003499
Dummy variable for age 41-50	-0.06007	0.003491
Dummy variable for age 51-60	-0.05909	0.004776
Dummy variable for age > 60	-0.06138	0.006789
Dummy variable for female	-0.06661	0.001721
mm1: Fail to yield, imp. turns, PXO, amber violations, etc	0.02736	0.001646
mm2: Disobey red lights; rail crossing violations.	0.045535	0.002791
mm3: Fail report accid.; Careless driving; Dang. driving;Crim. neg. caus. death	0.048755	0.004567
mm4: Unsafe move;imp. o'taking; Disobey signs	0.059365	0.003723
mm5: Fail to remain; breath test; alcohol; impairment.	0.068515	0.015328
mm6: Seat belt; F.T.C.; parking; divided h'way offences.	0.03421	0.001961
nn1: Minor neglect: no drivers license; permits; insurance, address change	0.031975	0.009674
nn2: Neglect: D/Lic. suspended or not produced; plates, insurance.	0.03312	0.003617
vv1: Minor vehicle neglect: lamps windows obstructed, etc.	0.051595	0.006479
vv2: Vehicle neglect: Unsafe veh., brakes, tires	0.08548	0.009013
vv3: Truck weight offences.	0.187235	0.025898
vv4: Truck dimension offences.	0.11554	0.011842
ee1: Environmental offences: noise, fumes.	0.091265	0.008340
mm7: Speeding offences	0.02902	0.000840

Appendix 3

Regression Run A2: Age, sex as dummies, tot. accs.

Description of variable	Estimated Coefficient	Standard Error
Intercept (Male, Age < 21)	0.1763	0.002876
Dummy variable for age 21-25	-0.03856	0.002852
Dummy variable for age 26-30	-0.0565	0.002913
Dummy variable for age 31-35	-0.06006	0.003044
Dummy variable for age 36-40	-0.05202	0.003516
Dummy variable for age 41-50	-0.06013	0.00351
Dummy variable for age 51-60	-0.05744	0.004832
Dummy variable for age > 60	-0.06156	0.007068
Dummy variable for female	-0.06122	0.001735
mm1: Fail to yield, imp. turns, PXO, amber violations, etc	0.02696	0.001753
mm2: Disobey red lights; rail crossing violations.	0.042125	0.002916
mm3: Fail report accid.; Careless driving; Dang. driving;Crim. neg. caus. death	0.02330	0.008613
mm4: Unsafe move;imp. o'taking; Disobey signs	0.06359	0.004959
mm5: Fail to remain; breath test; alcohol; impairment.	0.2444	0.054575
mm6: Seat belt; F.T.C.; parking; divided h'way offences.	0.2444	0.054575
nn1: Minor neglect: no drivers license; permits; insurance, address change	0.02624	0.010463
nn2: Neglect: D/Lic. suspended or not produced; plates, insurance.	0.03366	0.003708
vv1: Minor vehicle neglect: lamps windows obstructed, etc.	0.03366	0.003708
vv2: Vehicle neglect: Unsafe veh., brakes, tires	0.08673	0.009296
vv3: Insecure load.	0.15922	0.027210
vv4: Weight and dimension offences.	0.11074	0.011895
ee1: Environmental offences: noise, fumes.	0.08748	0.008334
mm7: Speeding offences	0.026515	0.000866
Total accidents in period 1	0.05831	0.001580

Appendix 3

Regression Run A5: Age, sex as dummies, fault accs.

Description of variable	Estimated Coefficient	Standard Error
Intercept (Male, Age < 21)	0.1811	0.002923
Dummy variable for age 21-25	-0.03879	0.002905
Dummy variable for age 26-30	-0.05755	0.002968
Dummy variable for age 31-35	-0.06119	0.003101
Dummy variable for age 36-40	-0.05309	0.003581
Dummy variable for age 41-50	-0.06141	0.003576
Dummy variable for age 51-60	-0.05906	0.004922
Dummy variable for age > 60	-0.06341	0.007197
Dummy variable for female	-0.06346	0.001764
mm1: Fail to yield, imp. turns, PXO, amber violations, etc	0.02833	0.001937
mm2: Disobey red lights; rail crossing violations.	0.044105	0.003221
mm3: Fail report accid.; Careless driving; Dang. driving;Crim. neg. caus. death	0.025255	0.009526
mm4: Unsafe move;imp. o'taking; Disobey signs	0.065445	0.005480
mm5: Fail to remain; breath test; alcohol; impairment.	0.2462	0.061209
mm6: Seat belt; F.T.C.; parking; divided h'way offences.	0.03359	0.002356
nn1: Minor neglect: no drivers license; permits; insurance, address change	0.02664	0.011584
nn2: Neglect: D/Lic. suspended or not produced; plates, insurance.	0.03417	0.004099
vv1: Minor vehicle neglect: lamps windows obstructed, etc.	0.05188	0.007244
vv2: Vehicle neglect: Unsafe veh., brakes, tires	0.08958	0.010293
vv3: Insecure load	0.166275	0.029826
vv4: Weight & dimension offences.	0.114335	0.013051
ee1: Environmental offences: noise, fumes.	0.089395	0.009204
mm7: Speeding offences	0.027835	0.000957
Total at fault accidents in period 1	0.056635	0.002495

Appendix 3

Regression Run A4: Age/sex as dummies, fault, not at fault accs.

Description of variable	Estimated Coefficient	Standard Error
Intercept (Male, Age < 21)	0.1763	0.002876
Dummy variable for age 21-25	-0.03856	0.002852
Dummy variable for age 26-30	-0.0565	0.002913
Dummy variable for age 31-35	-0.06006	0.003044
Dummy variable for age 36-40	-0.05202	0.003516
Dummy variable for age 41-50	-0.06013	0.00351
Dummy variable for age 51-60	-0.05744	0.004832
Dummy variable for age > 60	-0.06156	0.007068
Dummy variable for female	-0.06122	0.001735
mm1: Fail to yield, imp. turns, PXO, amber violations, etc	0.02697	0.001900
mm2: Disobey red lights; rail crossing violations.	0.04211	0.003158
mm3: Fail report accid.; Careless driving; Dang. driving;Crim. neg. caus. death	0.02344	0.009338
mm4: Unsafe move;imp. o'taking; Disobey signs	0.06367	0.005372
mm5: Fail to remain; breath test; alcohol; impairment.	0.2443	0.060010
mm6: Seat belt; F.T.C.; parking; divided h'way offences.	0.03222	0.002310
nn1: Minor neglect: no drivers license; permits; insurance, address change	0.02638	0.011354
nn2: Neglect: D/Lic. suspended or not produced; plates, insurance.	0.03374	0.004018
vv1: Minor vehicle neglect: lamps windows obstructed, etc.	0.04954	0.007103
vv2: Vehicle neglect: Unsafe veh., brakes, tires	0.08675	0.010091
vv3: Insecure load	0.15859	0.029246
vv4: Weight & dimension offences.	0.1109	0.012793
ee1: Environmental offences: noise, fumes.	0.08756	0.009020
mm7: Speeding offences	0.02650	0.000939
Total at fault accidents in period 1	0.05451	0.002446
Total driving properly accidents, period 1	0.06226	0.002484

Appendix 3

Regression Run B1: No Age/sex variables, no accs.

Description of variable	Estimated Coefficient	Standard Error
Intercept	0.1083	0.001486
mm1: Fail to yield, imp. turns, PXO, amber violations, etc	0.03224	0.001790
mm2: Disobey red lights; rail crossing violations.	0.05341	0.003035
mm3: Fail report accid.; Careless driving; Dang. driving;Crim. neg. caus. death	0.06617	0.004958
mm4: Unsafe move;imp. o'taking; Disobey signs	0.06722	0.004054
mm5: Fail to remain; breath test; alcohol; impairment.	0.079245	0.016679
mm6: Seat belt; F.T.C.; parking; divided h'way offences.	0.043205	0.002128
nn1: Minor neglect: no drivers license; permits; insurance, address change	0.043055	0.010544
nn2: Neglect: D/Lic. suspended or not produced; plates, insurance.	0.04907	0.003926
vv1: Minor vehicle neglect: lamps windows obstructed, etc.	0.06633	0.007045
vv2: Vehicle neglect: Unsafe veh., brakes, tires	0.10202	0.009826
vv3: Insecure load	0.2004	0.027991
vv4: Weight & dimension offences.	0.12488	0.012789
ee1: Environmental offences: noise, fumes.	0.1207	0.009048
mm7: Speeding offences	0.03416	0.000907

Appendix 3

Regression Run B2: No Age/sex variables, total accs.

Description of variable	Estimated Coefficient	Standard Error
Intercept	0.10335	0.001520
mm1: Fail to yield, imp. turns, PXO, amber violations, etc	0.03148	0.001886
mm2: Disobey red lights; rail crossing violations.	0.04925	0.003138
mm3: Fail report accid.; Careless driving; Dang. driving;Crim. neg. caus. death	0.041585	0.009278
mm4: Unsafe move;imp. o'taking; Disobey signs	0.072585	0.005341
mm5: Fail to remain; breath test; alcohol; impairment.	0.25755	0.059830
mm6: Seat belt; F.T.C.; parking; divided h'way offences.	0.03992	0.002290
nn1: Minor neglect: no drivers license; permits; insurance, address change	0.035915	0.011297
nn2: Neglect: D/Lic. suspended or not produced; plates, insurance.	0.04807	0.003981
vv1: Minor vehicle neglect: lamps windows obstructed, etc.	0.062545	0.007060
vv2: Vehicle neglect: Unsafe veh., brakes, tires	0.101415	0.010035
vv3: Insecure load	0.169635	0.029073
vv4: Weight & dimension offences.	0.11832	0.012712
ee1: Environmental offences: noise, fumes.	0.114985	0.008946
mm7: Speeding offences	0.03063	0.000927
Total accidents in period 1	0.065045	0.001696

Appendix 3

Regression Run B3: No Age/sex variables, fault accs.

Description of variable	Estimated Coefficient	Standard Error
Intercept	0.10695	0.001520
mm1: Fail to yield, imp. turns, PXO, amber violations, etc	0.03314	0.001893
mm2: Disobey red lights; rail crossing violations.	0.051685	0.003150
mm3: Fail report accid.; Careless driving; Dang. driving;Crim. neg. caus. death	0.04427	0.009317
mm4: Unsafe move;imp. o'taking; Disobey signs	0.074925	0.005363
mm5: Fail to remain; breath test; alcohol; impairment.	0.2602	0.059893
mm6: Seat belt; F.T.C.; parking; divided h'way offences.	0.04174	0.002298
nn1: Minor neglect: no drivers license; permits; insurance, address change	0.036755	0.011343
nn2: Neglect: D/Lic. suspended or not produced; plates, insurance.	0.04918	0.003997
vv1: Minor vehicle neglect: lamps windows obstructed, etc.	0.06561	0.007088
vv2: Vehicle neglect: Unsafe veh., brakes, tires	0.10512	0.010074
vv3: Insecure load	0.178085	0.029239
vv4: Weight & dimension offences.	0.122515	0.012786
ee1: Environmental offences: noise, fumes.	0.118	0.008981
mm7: Speeding offences	0.03226	0.000929
Total at fault accidents in period 1	0.06507	0.002433

Appendix 3

Regression Run B4: No Age/sex var., fau.,n/f. accs.

Description of variable	Estimated Coefficient	Standard Error
Intercept	0.10335	0.001494
mm1: Fail to yield, imp. turns, PXO, amber violations, etc	0.031495	0.001855
mm2: Disobey red lights; rail crossing violations.	0.049245	0.003085
mm3: Fail report accid.; Careless driving; Dang. driving;Crim. neg. caus. death	0.0417	0.009119
mm4: Unsafe move;imp. o'taking; Disobey signs	0.07265	0.005251
mm5: Fail to remain; breath test; alcohol; impairment.	0.25755	0.058638
mm6: Seat belt; F.T.C.; parking; divided h'way offences.	0.039935	0.002251
nn1: Minor neglect: no drivers license; permits; insurance, address change	0.03602	0.011103
nn2: Neglect: D/Lic. suspended or not produced; plates, insurance.	0.048135	0.003913
vv1: Minor vehicle neglect: lamps windows obstructed, etc.	0.06255	0.006942
vv2: Vehicle neglect: Unsafe veh., brakes, tires	0.101415	0.009862
vv3: Insecure load	0.16913	0.028628
vv4: Weight & dimension offences.	0.11844	0.012517
ee1: Environmental offences: noise, fumes.	0.115085	0.008794
mm7: Speeding offences	0.030625	0.000911
Total at fault accidents in period 1	0.06248	0.002384
Total driving properly accidents, period 1	0.06775	0.002427

Appendix 3

Regression Run C1: Age,sex; total cons., no accs.

Description of variable	Estimated Coefficient	Standard Error
Intercept (Male, Age < 21)	0.1839	0.00284
Dummy variable for age 21-25	-0.03752	0.002832
Dummy variable for age 26-30	-0.0564	0.002894
Dummy variable for age 31-35	-0.0709	0.003025
Dummy variable for age 36-40	-0.05246	0.003499
Dummy variable for age 41-50	-0.06007	0.003491
Dummy variable for age 51-60	-0.05909	0.004776
Dummy variable for age > 60	-0.06138	0.006789
Dummy variable for female	-0.06661	0.001721
Total convictions, period 1	0.03212	0.000812

Appendix 3

Regression Run C2: Age,sex; total cons., total accs.

Description of variable	Estimated Coefficient	Standard Error
Intercept (Male, Age < 21)	0.1763	0.002876
Dummy variable for age 21-25	-0.03856	0.002852
Dummy variable for age 26-30	-0.0565	0.002913
Dummy variable for age 31-35	-0.06006	0.003044
Dummy variable for age 36-40	-0.05202	0.003516
Dummy variable for age 41-50	-0.06013	0.00351
Dummy variable for age 51-60	-0.05744	0.004832
Dummy variable for age > 60	-0.06156	0.007068
Dummy variable for female	-0.06122	0.001735
Total convictions, period 1	0.02951	0.000827
Total Accidents, period 1	0.05902	0.001674

Appendix 3

Regression Run C3: Age,sex; total cons., fault accs.

Description of variable	Estimated Coefficient	Standard Error
Intercept (Male, Age < 21)	0.1811	0.002923
Dummy variable for age 21-25	-0.03879	0.002905
Dummy variable for age 26-30	-0.05755	0.002968
Dummy variable for age 31-35	-0.06119	0.003101
Dummy variable for age 36-40	-0.05309	0.003581
Dummy variable for age 41-50	-0.06141	0.003576
Dummy variable for age 51-60	-0.05906	0.004922
Dummy variable for age > 60	-0.06341	0.007197
Dummy variable for female	-0.06346	0.001764
Total convictions, period 1	0.03102	0.000840
At Fault Accidents, period 1	0.05743	0.002434

Appendix 3

Regression Run C4: Age,sex; tot. cons., f.,n/f. accs.

Description of variable	Estimated Coefficient	Standard Error
Intercept (Male, Age < 21)	0.1763	0.002876
Dummy variable for age 21-25	-0.03856	0.002852
Dummy variable for age 26-30	-0.0565	0.002913
Dummy variable for age 31-35	-0.06006	0.003044
Dummy variable for age 36-40	-0.05202	0.003516
Dummy variable for age 41-50	-0.06013	0.00351
Dummy variable for age 51-60	-0.05744	0.004832
Dummy variable for age > 60	-0.06156	0.007068
Dummy variable for female	-0.06122	0.001735
Total convictions, period 1	0.02951	0.000827
At Fault Accidents, period 1	0.05526	0.002388
Driving properly accidents, period 1	0.06289	0.002426

Appendix 3

Regression Run D1: No age,sex; tot. cons., no accs.

Description of variable	Estimated Coefficient	Standard Error
Intercept	0.1073	0.001424
Total convictions, period 1	0.03891	0.000773

Appendix 3

Regression Run D2: No age,sex; tot. cons., tot. accs.

Description of variable	Estimated Coefficient	Standard Error
Intercept	0.1019	0.001431
Total convictions, period 1	0.03518	0.000785
Total Accidents, period 1	0.06618	0.001604

Appendix 3

Regression Run D3: No age,sex; tot. cons., fau. accs.

Description of variable	Estimated Coefficient	Standard Error
Intercept	0.1053	0.001454
Total convictions, period 1	0.03705	0.000798
At fault Accidents, period 1	0.06644	0.002338

Appendix 3

Regression Run D4: No age,sex; tot.cons., f.n/f. accs.

Description of variable	Estimated Coefficient	Standard Error
Intercept	0.1019	0.001431
Total convictions, period 1	0.03518	0.000785
At fault Accidents, period 1	0.06378	0.002291
Driving properly accidents, period 1	0.06862	0.002334

APPENDIX 4

Theoretical Framework

Theorectical Framework

Each driver has an expected number of accidents for 1983-1984. For driver i this will be denoted by m_i. Consider a reference population of drivers identical in gender, age-group, and also in their 1981-1982 count of convictions and accidents. The average of their m's is E{m}. Since not all the m's in the reference population are the same, there is also a variance of m's, VAR{m}.

We estimate E{m} by

$$E\{m\} = a_1 + a_g + a_s + b_1x_1 + b_2x_2 + \ldots \qquad (1)$$

where

a_1 is a parameter for the "Base Driver" (who is a male in the 17-21 age group).

a_g is a parameter, g = 2,3,...8, depending on the age group as shown below

2	21-25 years old in 1982	
3	26-30	" " " "
4	31-35	" " " "
5	36-40	" " " "
6	41-50	" " " "
7	51-60	" " " "
8	> 60	" " " "

a_s is a parameter for "female" and b_1, b_2 .. are the "weights or "parameters" for variables x_1, x_2 etc., which are the numbers of convictions or accidents of a given type during 1981-1982.

We estimate the variance of the m's by

$$VAR\{m\} = [E\{m\}]^2/k \qquad (2)$$

where k is a parameter.

A4.2 Estimation of Parameters

We have estimated the parameters of equation (1) by a software well suited for accident counts. It facilitates the appropriate representation of accident counts by the negative binomial distribution and yields maximum likelihood estimates for the parameters. (Baker, R.J. and Nelder, J.A., The GLIM system. Generalized Linear Interactive Modelling, 1978.)

The parameter k in equation (2) has also been estimated by maximum likelihood methods. A special program had to be written for this purpose.

A4.3 Estimation of E{m} and VAR{m} from Accident Counts

It can be shown that if for every driver the accident occurrence obeys the Poisson Probability Law, then, for a group of drivers,

$E\{m\} = x$ = sample mean of accident counts

$VAR\{m\} = s^2 - x$ where s^2 is the sample variance of accident counts.

A4.4 The "Gamma Assumption"

Much of the analysis in section 4 rests on the assumption that the distribution of m's (accident potential) in the population of the Ontario drivers can be represented by a Gamma probability distribution function. As noted in the main text in section 4.4 we cannot be sure this assumption is a valid one until we check whether the accident counts which are observed support the "Gamma Assumption".

In tables A4.1 to A4.9 we check whether the number of drivers observed to have 0,1,2,... accidents can be well reproduced if one uses only x, s^2, and the Poisson and Gamma assumptions. Thus, for example, there were 6923 males, 17-20 years old, who during 1981-1984 had 1 accident (see Table A4.1). During the same period of time for males who were 17-20 years old, the sample mean was 0.463 accidents and the sample standard deviation was 0.718 accidents. Using these two sample statistics, we should expect 6867 drivers to have 1 accident if the "Gamma Assumption" was correct (Table A4.1). What has been observed and what is expected if the Gamma Assumption is correct seems to correspond. In fact, all pairs of numbers in all tables display good correspondence.

In tables A4.1 - A4.9 the Gamma assumption was tested "weakly". We used two numbers (x,s^2) to generate 5 to 9 other numbers and showed that these matched well what has been observed. We now proceed with a stronger test.

In tables A4.10 to A4.23 we examine 14 two period tables. Thus, for example, in Table A4.10 we show that 3214 male drivers, 17-20 years of age, had 0 accidents in 1981-1982 and 1 accident during 1983-1984. Now, using only x,s^2 for 1981-1982 and x for 1983-1984, we use the Gamma assumption to estimate what would be the expected number of such drivers - 3232.3. Once again, the correspondence in all 15 tables is impressive.

We conclude that the information we have about accident counts for Ontario drivers does not contradict the appropriateness of making use of the Gamma Assumption.

Appendix 4

ABLE A4.1: ACCIDENT DISTRIBUTION FOR 17-20 YEAR OLDS

No. of accs 1981-84 (x)	MALES (x=.463, s=.718) Drivers with x accs. Observed	Estimated	FEMALES (x=.192, s=.452) Drivers with x accs. Observed	Estimated
0	16507	16537	16966	16977
1	6923	6867	3092	3060
2	1775	1782	339	369
3	343	370	45	37
4	75	67	5	3
5	13	11		
6	0	2		
All	25636		20447	

TABLE A4.2: ACCIDENT DISTRIBUTION FOR 21-25 YEAR OLDS

No. of accs 1981-84 (x)	MALES (x=.379, s=.664) Drivers with x accs. Observed	Estimated	FEMALES (x=.164, s=.423) Drivers with x accs. Observed	Estimated
0	45054	45224	50331	50359
1	15182	14778	7623	7557
2	3203	3441	855	889
3	662	693	84	95
4	154	129	11	10
5	28	23	7	1
6	5	4		
7	4	1		
All	64292		58911	

Appendix 4

TABLE A4.3: ACCIDENT DISTRIBUTION FOR 26-30 YEAR OLDS

	MALES (x=.306, s=.609)		FEMALES (x=.132, s=.380)	
No. of accs 1981-84 (x)	Drivers with x accs. Observed	Estimated	Drivers with x accs. Observed	Estimated
0	49857	50016	57888	57902
1	12961	12607	6968	6946
2	2528	2700	734	726
3	493	544	52	72
4	120	106	7	7
5	23	20	4	1
6	12	4	0	0
7	1	1	1	0
8	2	0		
9	1	0		
All	65998		65654	

TABLE A4.4: ACCIDENT DISTRIBUTION FOR 31-35 YEAR OLDS

	MALES (x=.289, s=.586)		FEMALES (x=.132, s=.382)	
No. of accs 1981-84 (x)	Drivers with x accs. Observed	Estimated	Drivers with x accs. Observed	Estimated
0	39864	39993	50866	50891
1	10005	9738	6177	6108
2	1861	1952	589	648
3	291	363	78	66
4	74	65	9	7
5	16	11	0	1
6	7	2	1	0
7	4	0		
8	2	0		
All	52124		57720	

Appendix 4

TABLE A4.5: ACCIDENT DISTRIBUTION FOR 36-40 YEAR OLDS

	MALES (x=.286, s=.594)		FEMALES (x=.137, s=.390)	
No. of accs 1981-84 (x)	Drivers with x accs. Observed	Estimated	Drivers with x accs. Observed	Estimated
0	20440	20502	29255	29254
1	4884	4744	3602	3609
2	927	1000	416	402
3	192	204	32	43
4	43	41	7	5
5	9	8	1	1
6	1	2		
7	5	0		
All	26501		33313	

TABLE A4.6: ACCIDENT DISTRIBUTION FOR 41-50 YEAR OLDS

	MALES (x=.261, s=.565)		FEMALES (x=.122, s=.366)	
No. of accs 1981-84 (x)	Drivers with x accs. Observed	Estimated	Drivers with x accs. Observed	Estimated
0	24587	24665	31084	31094
1	5420	5256	3460	3435
2	970	1041	336	348
3	186	201	27	34
4	28	38	8	3
5	10	7	1	0
6	3	1		
7	5	0		
8	1	0		
All	31210		34916	

Appendix 4

TABLE A4.7: ACCIDENT DISTRIBUTION FOR 51-60 YEAR OLDS

	MALES (x=.229, s=.518)		FEMALES (x=.108, s=.350)	
No. of accs 1981-84 (x)	Drivers with x accs. Observed	Estimated	Drivers with x accs. Observed	Estimated
0	12119	12113	14873	14864
1	2361	2375	1387	1410
2	413	406	170	152
3	68	66	16	17
4	10	10	0	2
5	1	2		
All	14972		16446	

TABLE A4.8: ACCIDENT DISTRIBUTION FOR > 60-YEAR OLDS

	MALES (x=.222, s=.521)		FEMALES (x=.117, s=.350)	
No. of accs 1981-84 (x)	Drivers with x accs. Observed	Estimated	Drivers with x accs. Observed	Estimated
0	6456	6469	7112	7111
1	1199	1173	788	792
2	197	213	66	63
3	41	39	4	4
4	5	7		
5	4	1		
All	7902		7970	

Appendix 4

TABLE A4.9: ACCIDENT DISTRIBUTION FOR ALL DRIVERS

No. of accs 1981-84 (x)	MALES (x=.32Ø, s=.618) Drivers with x accs. Observed	Estimated	FEMALES (x=.14Ø, s=.393) Drivers with x accs. Observed	Estimated
Ø	214884	215424	258375	258450
1	58935	57736	33Ø97	32924
2	11874	12441	35Ø5	3592
3	2276	2463	338	37Ø
4	5Ø9	466	47	37
5	1Ø4	86	13	4
6	28	16	1	Ø
7	19	3	1	Ø
8	5	1		
9	1	Ø		
All	288635		295377	

Appendix 4

TABLE A4.10: 2-PERIOD ACCIDENT FREQUENCIES FOR MALES AGED 17-20 YEARS
(Period 1: Mean=0.2461, Var(x)=0.2605; Period 2: Mean=0.2173)

Per. 1 Accs.	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated
	Observed and estimated no. of drivers with period 2 accident counts =											
	0		1		2		3		4		5	
0	16507	16519.8	3214	3232.3	422	391.6	37	37.7	5	3.2	0	0.2
1	3709	3660.5	840	886.9	122	128.1	15	14.3	3	1.4	0	0.1
2	513	502.2	125	145.1	31	24.3	6	3.1	0	0.3	0	0.0
3	59	54.8	19	18.4	2	3.5	0	0.5	0	0.1	0	0.0
4	5	5.2	1	2.0	0	0.4	0	0.1	0	0.0	0	0.0
5	1	0.5	0	0.2	0	0.1	0	0.0	0	0.0	0	0.0

TABLE A4.11: 2-PERIOD ACCIDENT FREQUENCIES FOR FEMALES AGED 17-20 YEARS
(Period 1: Mean=0.1057, Var(x)=0.1085; Period 2: Mean=0.0886)

Per. 1 Accs.	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated
	Observed and estimated no. of drivers with period 2 accident counts =											
	0		1		2		3		4		5	
0	16966	16951.9	1382	1394.9	70	71.9	3	3.0	0	0.1	0	0.0
1	1710	1708.9	173	176.1	15	10.9	1	0.5	0	0.0	0	0.0
2	96	107.9	20	13.4	3	1.0	0	0.1	0	0.0	0	0.0
3	7	5.5	1	0.8	0	0.1	0	0.0	0	0.0	0	0.0
4	0	0.2	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
5	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0

Appendix 4

TABLE A4.12: 2-PERIOD ACCIDENT FREQUENCIES FOR MALES AGED 21-25 YEARS
(Period 1: Mean=0.2049, Var(x)=0.2231; Period 2: Mean=0.1746)

Per. 1 Accs.	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated
	Observed and estimated no. of drivers with period 2 accident counts =											
	0		1		2		3		4		5	
0	45054	45249.8	6917	6781.8	732	729.1	72	68.1	14	5.9	1	0.5
1	8265	7958.7	1503	1711.2	246	239.7	39	27.6	2	2.8	1	0.3
2	968	1004.1	224	281.3	49	48.6	6	6.6	3	0.8	0	0.1
3	120	110.0	37	38.0	8	7.8	0	1.2	0	0.2	0	0.0
4	15	11.1	8	4.6	1	1.1	1	0.2	0	0.0	0	0.0
5	3	1.1	0	0.5	1	0.1	0	0.0	0	0.0	0	0.0

TABLE A4.13: 2-PERIOD ACCIDENT FREQUENCIES FOR FEMALES AGED 21-25 YEARS
(Period 1: Mean=0.0861, Var(x)=0.0901; Period 2: Mean=0.0779)

Per. 1 Accs.	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated
	Observed and estimated no. of drivers with period 2 accident counts =											
	0		1		2		3		4		5	
0	50331	50337.8	3592	3606.2	219	198.0	5	9.8	1	0.5	1	0.0
1	4031	3988.1	410	437.8	24	32.4	1	2.0	4	0.1	0	0.0
2	226	242.1	39	35.8	6	3.3	1	0.2	0	0.0	0	0.0
3	16	13.2	3	2.5	0	0.3	0	0.0	0	0.0	0	0.0
4	0	0.7	1	0.2	0	0.0	0	0.0	0	0.0	0	0.0
5	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0

Appendix 4

TABLE A4.14: 2-PERIOD ACCIDENT FREQUENCIES FOR MALES AGED 26-30 YEARS
(Period 1: Mean=0.1608, Var(x)=0.1800; Period 2: Mean=0.1452)

Per. 1 Accs.	Observed and estimated no. of drivers with period 2 accident counts = 0		1		2		3		4		5	
0	49857	50094.0	6121	5926.5	615	611.0	52	59.9	11	5.7	0	0.5
1	6840	6562.8	1205	1353.3	165	199.0	28	25.3	4	3.0	1	0.3
2	708	749.3	203	220.4	39	42.1	10	6.6	3	0.9	0	0.1
3	73	81.3	32	31.1	3	7.3	1	1.4	0	0.2	0	0.0
4	10	8.6	4	4.0	7	1.1	0	0.2	0	0.0	0	0.0
5	2	0.9	0	0.5	1	0.2	1	0.0	1	0.0	0	0.0

TABLE A4.15: 2-PERIOD ACCIDENT FREQUENCIES FOR FEMALES AGED 26-30 YEARS
(Period 1: Mean=0.0658, Var(x)=0.0692; Period 2: Mean=0.0659)

Per. 1 Accs.	Observed and estimated no. of drivers with period 2 accident counts = 0		1		2		3		4		5	
0	57888	57922.2	3463	3460.2	213	185.0	8	9.5	1	0.5	0	0.0
1	3505	3449.8	340	368.8	13	28.4	2	1.9	1	0.1	0	0.0
2	181	183.9	23	28.3	0	2.9	2	0.2	0	0.0	0	0.0
3	8	9.4	3	1.9	1	0.2	0	0.0	0	0.0	0	0.0
4	1	0.5	0	0.1	0	0.0	1	0.0	0	0.0	0	0.0
5	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0

Appendix 4

TABLE A4.16: 2-PERIOD ACCIDENT FREQUENCIES FOR MALES AGED 31-35 YEARS
(Period 1: Mean=Ø.1495, Var(x)=Ø.1635; Period 2: Mean=Ø.1394)

Per. 1 Accs.	Observed and estimated no. of drivers with period 2 accident counts = Ø		1		2		3		4		5	
Ø	39864	39963.9	48Ø3	4719.6	443	452.8	35	4Ø.1	2	3.4	1	Ø.3
1	52Ø2	5Ø62.6	934	971.4	1ØØ	129.Ø	21	14.6	3	1.5	Ø	Ø.1
2	484	521.Ø	114	138.4	24	23.5	4	3.2	3	Ø.4	Ø	Ø.Ø
3	42	49.5	19	16.8	5	3.5	3	Ø.6	Ø	Ø.1	1	Ø.Ø
4	8	4.5	2	1.9	1	Ø.5	1	Ø.1	1	Ø.Ø	Ø	Ø.Ø
5	1	Ø.4	Ø	Ø.2	2	Ø.1	Ø	Ø.Ø	Ø	Ø.Ø	Ø	Ø.Ø

TABLE A4.17: 2-PERIOD ACCIDENT FREQUENCIES FOR FEMALES AGED 31-35 YEARS
(Period 1: Mean=Ø.Ø635, Var(x)=Ø.Ø673; Period 2: Mean=Ø.Ø687)

Per. 1 Accs.	Observed and estimated no. of drivers with period 2 accident counts = Ø		1		2		3		4		5	
Ø	5Ø866	5Ø95Ø.5	3216	3118.1	166	183.4	11	1Ø.6	Ø	Ø.6	Ø	Ø.Ø
1	2961	2886.Ø	281	339.5	25	29.5	3	2.3	Ø	Ø.2	Ø	Ø.Ø
2	142	157.1	31	27.3	3	3.2	Ø	Ø.3	Ø	Ø.Ø	Ø	Ø.Ø
3	11	8.4	2	1.9	Ø	Ø.3	1	Ø.Ø	Ø	Ø.Ø	Ø	Ø.Ø
4	1	Ø.4	Ø	Ø.1	Ø	Ø.Ø	Ø	Ø.Ø	Ø	Ø.Ø	Ø	Ø.Ø
5	Ø	Ø.Ø	Ø	Ø.Ø	Ø	Ø.Ø	Ø	Ø.Ø	Ø	Ø.Ø	Ø	Ø.Ø

TABLE A4.18: 2-PERIOD ACCIDENT FREQUENCIES FOR MALES AGED 36-40 YEARS
(Period 1: Mean=0.1476, Var(x)=0.1641; Period 2: Mean=0.1381)

Per. 1 Accs.	Observed and estimated no. of drivers with period 2 accident counts = 0		1		2		3		4		5	
0	20440	20460.3	2339	2323.8	227	231.7	17	22.0	2	2.0	0	0.2
1	2545	2483.8	487	495.4	64	70.7	9	8.7	0	1.0	0	0.1
2	213	264.7	77	75.5	20	14.0	1	2.1	0	0.3	0	0.0
3	34	26.9	12	10.0	6	2.3	1	0.4	3	0.1	0	0.0
4	0	2.7	2	1.2	0	0.3	1	0.1	0	0.0	0	0.0
5	0	0.3	0	0.1	1	0.0	0	0.0	0	0.0	0	0.0

TABLE A4.19: 2-PERIOD ACCIDENT FREQUENCIES FOR FEMALES AGED 36-40 YEARS
(Period 1: Mean=0.0699, Var(x)=0.0732; Period 2: Mean=0.0671)

Per. 1 Accs.	Observed and estimated no. of drivers with period 2 accident counts = 0		1		2		3		4		5	
0	29255	29224.5	1757	1792.2	100	92.5	2	4.5	1	0.2	0	0.0
1	1845	1868.4	218	192.8	7	14.0	3	0.9	0	0.1	0	0.0
2	98	100.5	18	14.6	2	1.4	1	0.1	0	0.0	0	0.0
3	5	5.1	1	0.9	0	0.1	0	0.0	0	0.0	0	0.0
4	0	0.2	0	0.1	0	0.0	0	0.0	0	0.0	0	0.0
5	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0

Appendix 4

TABLE A4.20: 2-PERIOD ACCIDENT FREQUENCIES FOR MALES AGED 41-50 YEARS
(Period 1: Mean=0.1324, Var(x)=0.1505; Period 2: Mean=0.1285)

Per. 1 Accs.	Observed and estimated no. of drivers with period 2 accident counts = 0		1		2		3		4		5	
0	24587	24773.9	2680	2507.1	238	258.0	26	26.7	4	2.8	0	0.3
1	2740	2582.7	470	531.5	66	82.5	7	11.4	3	1.5	1	0.2
2	262	273.8	70	85.0	7	17.6	1	3.1	0	0.5	0	0.1
3	24	29.2	9	12.1	4	3.1	0	0.7	1	0.1	0	0.0
4	1	3.1	1	1.6	0	0.5	0	0.1	0	0.0	0	0.0
5	1	0.3	2	0.2	3	0.1	0	0.0	0	0.0	0	0.0

TABLE A4.21: 2-PERIOD ACCIDENT FREQUENCIES FOR FEMALES AGED 41-50 YEARS
(Period 1: Mean=0.0624, Var(x)=0.0664; Period 2: Mean=0.0593)

Per. 1 Accs.	Observed and estimated no. of drivers with period 2 accident counts = 0		1		2		3		4		5	
0	31084	31136.3	1677	1639.0	97	88.3	0	4.8	1	0.3	0	0.0
1	1783	1722.8	155	185.6	10	15.1	2	1.1	0	0.1	0	0.0
2	84	97.5	10	15.9	2	1.7	0	0.2	0	0.0	0	0.0
3	7	5.6	2	1.2	0	0.2	0	0.0	0	0.0	0	0.0
4	1	0.3	0	0.1	0	0.0	0	0.0	0	0.0	0	0.0
5	1	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0

Appendix 4

TABLE A4.22: 2-PERIOD ACCIDENT FREQUENCIES FOR MALES AGED 51-60 YEARS
(Period 1: Mean=0.1196, Var(x)=0.1264; Period 2: Mean=0.1099)

Per. 1 Accs.	Observed and estimated no. of drivers with period 2 accident counts = 0		1		2		3		4		5	
0	12119	12041.9	1108	1193.3	101	87.3	6	5.6	0	0.3	0	0.0
1	1253	1297.7	206	189.9	26	18.4	7	1.5	0	0.1	0	0.0
2	106	103.3	32	20.0	2	2.4	0	0.2	0	0.0	0	0.0
3	4	7.2	1	1.7	1	0.3	0	0.0	0	0.0	0	0.0
4	0	0.5	0	0.1	0	0.0	0	0.0	0	0.0	0	0.0
5	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0

TABLE A4.23: 2-PERIOD ACCIDENT FREQUENCIES FOR FEMALES AGED 51-60 YEARS
(Period 1: Mean=0.0557, Var(x)=0..0598; Period 2: Mean=0.0522)

Per. 1 Accs.	Observed and estimated no. of drivers with period 2 accident counts = 0		1		2		3		4		5	
0	14873	14867.0	672	680.2	42	36.0	1	2.0	0	0.1	0	0.0
1	715	725.3	76	76.7	10	6.4	0	0.5	0	0.0	0	0.0
2	52	40.9	4	6.8	0	0.8	0	0.1	0	0.0	0	0.0
3	1	2.4	0	0.5	0	0.1	0	0.0	0	0.0	0	0.0
4	0	0.1	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
5	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0

TABLE A4.24: 2-PERIOD ACCIDENT FREQUENCIES FOR MALES AGED > 60 YEARS
(Period 1: Mean=0.1212, Var(x)=0.1392; Period 2: Mean=0.1006)

Per. 1 Accs.	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated
	0		1		2		3		4		5	
0	6456	6490.4	558	513.2	35	45.1	2	4.1	0	0.4	0	0.0
1	641	618.6	99	100.8	16	14.8	1	1.8	0	0.2	0	0.0
2	63	65.6	17	17.9	0	3.3	0	0.5	0	0.1	0	0.0
3	6	7.2	3	2.7	2	0.6	0	0.1	0	0.0	0	0.0
4	1	0.8	2	0.4	0	0.1	0	0.0	0	0.0	0	0.0
5	0	0.1	0	0.1	0	0.0	0	0.0	0	0.0	0	0.0

TABLE A4.25: 2-PERIOD ACCIDENT FREQUENCIES FOR FEMALES AGED > 60 YEARS
(Period 1: Mean=0.0572, Var(x)=0.0587; Period 2: Mean=0.0597)

Per. 1 Accs.	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated	Observed	Estimated
	0		1		2		3		4		5	
0	7112	7111.8	406	403.2	15	16.6	0	0.6	0	0.0	0	0.0
1	382	386.3	33	31.9	3	1.7	0	0.1	0	0.0	0	0.0
2	18	15.3	1	1.7	0	0.1	0	0.0	0	0.0	0	0.0
3	0	0.5	0	0.1	0	0.0	0	0.0	0	0.0	0	0.0
4	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
5	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0